



PCT/FR 00/01908

29 AOUT 2000

10/019965

BREVET D'INVENTION

FR 00/01908

G

CERTIFICAT D'UTILITÉ - CERTIFICAT D'ADDITION

COPIE OFFICIELLE

REC'D 18 SEP 2000

WIPO

PCT

Le Directeur général de l'Institut national de la propriété industrielle certifie que le document ci-annexé est la copie certifiée conforme d'une demande de titre de propriété industrielle déposée à l'Institut.

12 JULI 2000

Fait à Paris, le

**PRIORITY
DOCUMENT**
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

Pour le Directeur général de l'Institut
national de la propriété industrielle
Le Chef du Département des brevets

Martine PLANCHE

BEST AVAILABLE COPY

INSTITUT
NATIONAL DE
LA PROPRIÉTÉ
INDUSTRIELLE

SIEGE
26 bis, rue de Saint Petersburg
75800 PARIS Cédex 08
Téléphone : 01 53 04 53 04
Télécopie : 01 42 93 59 30

THIS PAGE BLANK (USPTO)

REQUÊTE EN DÉLIVRANCE

26 bis, rue de Saint Pétersbourg
75800 Paris Cedex 08
Téléphone : 01 53 04 53 04 Télécopie : 01 42 93 59 30

Confirmation d'un dépôt par télécopie ☐

Cet imprimé est à remplir à l'encre noire en lettres capitales

Réservé à l'INPI

DATE DE REMISE DES PIÈCES **5 JUIL 1999**
N° D'ENREGISTREMENT NATIONAL **9908634**
DÉPARTEMENT DE DÉPÔT **75 INPI PARIS**
DATE DE DÉPÔT **05 JUIL. 1999**

1 NOM ET ADRESSE DU DEMANDEUR OU DU MANDATAIRE
À QUI LA CORRESPONDANCE DOIT ÊTRE ADRESSÉE

CABINET PLASSERAUD
84, rue d'Amsterdam
75440 PARIS CEDEX 09

2 DEMANDE Nature du titre de propriété industrielle

☒ brevet d'invention ☐ demande divisionnaire

☐ certificat d'utilité ☐ transformation d'une demande de brevet européen

☐ demande tridimensionnelle

☐ brevet d'invention

n° du pouvoir permanent références du correspondant téléphone

BLO/FC-BFF990135 01 44 63 41 11

Établissement du rapport de recherche

☐ différé ☒ immédiat

Le demandeur, personne physique, requiert le paiement échelonné de la redevance ☐ oui ☐ non

Titre de l'invention (200 caractères maximum)

PROCEDE ET DISPOSITIF DE CODAGE AUDIO

3 DEMANDEUR (S) n° SIREN

code APE-NAF

Norm et prénoms (souligner le nom patronymique) ou dénomination

MATRA NORTEL COMMUNICATIONS

Forme juridique

**Société par Actions
Simplifiée**

Nationalité (s) **Française**

Adresse (s) complète (s)

50, rue du Président Sadate
29100 QUIMPER

Pays

FRANCE

En cas d'insuffisance de place, poursuivre sur papier libre ☐

4 INVENTEUR (S) Les inventeurs sont les demandeurs ☐ oui ☒ non

Si la réponse est non, fournir une désignation séparée

5 RÉDUCTION DU TAUX DES REDEVANCES

☐ requise pour la 1ère fois

☐ requise antérieurement au dépôt ; joindre copie de la décision d'admission

6 DÉCLARATION DE PRIORITÉ OU REQUÊTE DU BÉNÉFICE DE LA DATE DE DÉPÔT D'UNE DEMANDE ANTÉRIEURE

pays d'origine

numéro

date de dépôt

nature de la demande

7 DIVISIONS

antérieures à la présente demande n°

date

n°

date

8 SIGNATURE DU DEMANDEUR OU DU MANDATAIRE

(nom et qualité du signataire)

B. LOISEL

CPI n° 94_0311

SIGNATURE DU PRÉPOSÉ À LA RÉCEPTION

SIGNATURE APRÈS ENREGISTREMENT DE LA DEMANDE À L'INPI



DEPARTEMENT DES BREVETS

26bis, rue de Saint-Petersbourg
75800 Paris Cédex 08
Tél. : 01 53 04 53 04 - Télécopie : 01 42 93 59 30

BREVET D'INVENTION, CERTIFICAT D'UTILITE

BLO/FC-BFF990135

DÉSIGNATION DE L'INVENTEUR

(si le demandeur n'est pas l'inventeur ou l'unique inventeur)

N° D'ENREGISTREMENT NATIONAL

9908634

TITRE DE L'INVENTION : PROCÉDE ET DISPOSITIF DE CODAGE AUDIO

LA DEMANDERESSE : MATRA NORTEL COMMUNICATIONS
avant pour Mandataire

LE(S) SOUSSIGNÉ(S)

CABINET PLASSERAUD
84, rue d'Amsterdam
75440 PARIS CEDEX 09

DÉSIGNE(NT) EN TANT QU'INVENTEUR(S) (indiquer nom, prénoms, adresse et souligner le nom patronymique) :

CAPMAN, François

47, rue des Etats-Généraux
78000 VERSAILLES
FRANCE

MURGIA, Carlo

31, avenue de Saint-Cloud
78000 VERSAILLES
FRANCE

NOTA : A titre exceptionnel, le nom de l'inventeur peut être suivi de celui de la société à laquelle il appartient (société d'appartenance) lorsque celle-ci est différente de la société déposante ou titulaire.

Date et signature (s) du demandeur (s) ou du mandataire Paris, le 5 juillet 1999


B. LOISEL
CPI n° 94-0311

PROCEDE ET DISPOSITIF DE CODAGE AUDIO

La présente invention concerne le domaine du codage des signaux audio. Elle s'applique en particulier, mais non exclusivement, au codage de la parole, en bande étroite ou en bande élargie, dans diverses gammes de débit
5 de codage.

La conception d'un codec audio vise principalement à fournir un bon ~~compromis entre le débit du flux transmis par le codeur et la qualité du signal~~
audio que le décodeur est capable de reconstruire à partir de ce flux.

Dans cette optique, il a notamment été développé des familles de
10 codeurs reposant sur une analyse du signal audio dans le domaine spectral : le codeur estime une fréquence fondamentale du signal, représentant sa hauteur tonale (« pitch »), et l'analyse spectrale consiste à déterminer des paramètres représentant la structure harmonique du signal aux fréquences qui sont des multiples entiers de cette fréquence fondamentale. Une modélisation de la
15 composante non-harmonique, ou non-voisée, peut également être effectuée dans le domaine spectral. Les paramètres transmis au décodeur représentent typiquement le module du spectre des composantes voisées et non-voisées. Il s'y ajoute des informations représentant soit des décisions voisé/non-voisé relatives à différentes portions de spectre, soit des informations sur la
20 probabilité de voisement du signal, permettant au décodeur de déterminer dans quelles portions du spectre il doit utiliser la composante voisée ou la composante non-voisée.

Ces familles de codeurs comprennent les codeurs du type MBE (« Multi-Band Excitation »), ou encore les codeurs de type STC (« Sinusoidal Transform Coder »). A titre de référence, on peut citer les brevets US
25 4 856 068, 4 885 790, 4 937 873, 5 054 072, 5 081 681, 5 195 166, 5 216 747, 5 226 084, 5 226 108, 5 247 579, 5 473 727, 5 517 511, 5 630 011, 5 630 012, 5 649 050, 5 651 093, 5 664 051, 5 664 052, 5 684 926, 5 701 390, 5 715 365, 5 749 065, 5 752 222, 5 765 127, 5 774 837 et 5 890 108.

30 Un but de la présente invention est de permettre d'améliorer la modélisation du module du spectre du signal, dans un schéma de codage à analyse dans le domaine spectral.

L'invention propose ainsi un procédé de codage d'un signal audio, dans lequel on estime une fréquence fondamentale du signal audio, on
35 détermine un spectre du signal audio par une transformée dans le domaine

fréquentiel d'une trame du signal audio, et on inclut dans un flux numérique de sortie des données de codage d'une composante harmonique du signal audio, comprenant des données représentatives d'amplitudes spectrales associées à des fréquences multiples de la fréquence fondamentale. Selon l'invention,
5 l'amplitude spectrale associée à une desdites fréquences multiples de la fréquence fondamentale est un maximum local du module du spectre au voisinage de ladite fréquence multiple.

L'invention propose également un codeur audio comprenant des moyens pour la mise en œuvre du procédé ci-dessus.

10 D'autres particularités et avantages de la présente invention apparaîtront dans la description ci-après d'exemples de réalisation non limitatifs, en référence aux dessins annexés, dans lesquels :

- la figure 1 est un schéma synoptique d'un codeur audio selon l'invention ;
- les figures 2 et 3 sont des diagrammes illustrant la formation des trames
15 de signal audio dans le codeur de la figure 1 ;
- les figures 4 et 5 sont des graphiques montrant un exemple de spectre du signal audio et illustrant l'extraction des enveloppes supérieure et inférieure de ce spectre ;
- la figure 6 est un schéma synoptique d'un exemple de moyens de
20 quantification utilisables dans le codeur de la figure 1 ;
- la figure 7 est un schéma synoptique de moyens utilisables pour extraire des paramètres se rapportant à la phase de la composante non-harmonique dans une variante du codeur de la figure 1 ;
- la figure 8 est un schéma synoptique d'un décodeur audio correspondant
25 au codeur de la figure 1 ;
- la figure 9 est un organigramme d'un exemple de procédure de lissage de coefficients spectraux et d'extraction de phases minimales mise en œuvre dans le décodeur de la figure 8 ;
- la figure 10 est un schéma synoptique de modules d'analyse et de
30 mixage spectral de composantes harmonique et non-harmonique du signal audio ;
- les figures 11 à 13 sont des graphiques montrant des exemples de fonctions non-linéaires utilisables dans le module d'analyse de la figure 10 ;

- les figures 14 et 15 sont des diagrammes illustrant une façon de procéder à la synthèse temporelle des trames de signal dans le décodeur de la figure 8 ;
- les figures 16 et 17 sont des graphiques montrant des fonctions de fenêtrage utilisables dans la synthèse des trames suivant les figures 14 et 15 ;
- les figures 18 et 19 sont des schémas synoptiques de moyens d'interpolation utilisables dans une variante de réalisation du codeur et du décodeur ;
- la figure 20 est un schéma synoptique de moyens d'interpolation utilisables dans une autre variante de réalisation du codeur ;
- les figures 21 et 22 sont des diagrammes illustrant une autre façon de procéder à la synthèse temporelle des trames de signal dans le décodeur de la figure 8, à l'aide d'une interpolation de paramètres ;
- les figures 23 à 25 sont des schémas synoptiques de variantes de moyens de post-traitement des coefficients cepstraux représentant l'enveloppe supérieure du spectre du signal dans le codeur de la figure 1 ; et
- la figure 26 est un schéma synoptique partiel d'un décodeur associé à un codeur selon la figure 25.

Le codeur et le décodeur décrits ci-après sont des circuits numériques qui peuvent, comme il est usuel dans le domaine du traitement des signaux audio, être réalisés par programmation d'un processeur de signal numérique (DSP) ou d'un circuit intégré d'application spécifique (ASIC).

Le codeur audio représenté sur la figure 1 traite un signal audio d'entrée x qui, dans l'exemple non-limitatif considéré ci-après, est un signal de parole. Le signal x est disponible sous forme numérique, par exemple à une fréquence d'échantillonnage F_e de 8 kHz. Il est par exemple délivré par un convertisseur analogique-numérique traitant le signal de sortie amplifié d'un microphone. Le signal d'entrée x peut également être formé à partir d'une autre version, analogique ou numérique, codée ou non, du signal de parole.

Le codeur comprend un module 1 qui forme des trames successives de signal audio pour les différents traitements effectués, et un multiplexeur de sortie 6 qui délivre un flux de sortie Φ contenant pour chaque trame des jeux de paramètres de quantification à partir desquels un décodeur sera capable de

synthétiser une version décodée du signal audio.

La structure des trames est illustrée par les figures 2 et 3. Chaque trame 2 est composée d'un nombre N d'échantillons consécutifs du signal audio x . Les trames successives présentent des décalages temporels mutuels correspondant à M échantillons, de sorte que leur recouvrement est de $L = N - M$ échantillons du signal. Dans l'exemple considéré, où $N = 256$, $M = 160$ et $L = 96$, la durée des trames 2 est de $N/F_e = 32$ ms, et une trame est formée toutes les $M/F_e = 20$ ms.

De façon classique, le module 1 multiplie les échantillons de chaque trame 2 par une fonction de fenêtrage f_A , de préférence choisie pour ses bonnes propriétés spectrales. Les échantillons $x(i)$ de la trame étant numérotés de $i = 0$ à $i = N - 1$, la fenêtre d'analyse $f_A(i)$ peut ainsi être une fenêtre de Hamming, d'expression :

$$f_A(i) = 0,54 + 0,46 \cdot \cos\left(2\pi \frac{i - (N-1)/2}{N}\right) \quad (1)$$

ou une fenêtre de Hanning d'expression :

$$f_A(i) = \frac{1}{2} \left(1 + \cos\left(2\pi \frac{i - (N-1)/2}{N}\right) \right) \quad (2)$$

ou encore une fenêtre de Kaiser, d'expression :

$$f_A(i) = \frac{I_0\left(\alpha \sqrt{1 - \left(\frac{i - (N-1)/2}{N}\right)^2}\right)}{I_0(\alpha)} \quad (3)$$

où α est un coefficient par exemple égal à 6, et $I_0(\cdot)$ désigne la fonction de Bessel d'indice 0.

Le codeur de la figure 1 procède à une analyse du signal audio dans le domaine spectral. Il comporte un module 3 qui calcule la transformée de Fourier rapide (TFR) de chaque trame de signal. La trame de signal est mise en forme avant d'être soumise au module de TFR 3 : le module 1 lui adjoint $N = 256$ échantillons à zéro afin d'obtenir la résolution maximale de la transformée de Fourier, et il effectue d'autre part une permutation circulaire des $2N = 512$ échantillons afin de compenser les effets de phase résultant de la fenêtre d'analyse. Cette modification de la trame est illustrée par la figure 3. La trame dont on calcule la transformée de Fourier rapide sur $2N = 512$ points commence par les $N/2 = 128$ derniers échantillons pondérés de la trame, suivis

par les $N = 256$ échantillons à zéro, et se termine par les $N/2 = 128$ premiers échantillons pondérés de la trame.

Le module de TFR 3 obtient le spectre du signal pour chaque trame, dont le module et la phase sont respectivement notés $|X|$ et ϕ_X , ou $|X(i)|$ et $\phi_X(i)$ pour les index de fréquence $i = 0$ à $i = 2N-1$ (grâce à la symétrie de la transformée de Fourier et des trames, on peut se limiter aux valeurs pour $0 \leq i < N$).

Un détecteur de fréquence fondamentale 4 estime pour chaque trame de signal une valeur de la fréquence fondamentale F_0 . Le détecteur 4 peut appliquer toute méthode connue d'analyse du signal de parole de la trame pour estimer la fréquence fondamentale F_0 , par exemple une méthode basée sur la fonction d'autocorrélation ou la fonction AMDF, éventuellement précédée d'un module de blanchiment par prédiction linéaire. L'estimation peut également être effectuée dans le domaine spectral ou dans le domaine cepstral. Une autre possibilité est d'évaluer les intervalles de temps entre les ruptures consécutives du signal de parole attribuables à des fermetures de la glotte du locuteur intervenant pendant la durée de la trame. Des méthodes bien connues utilisables pour détecter de telles micro-ruptures sont décrites dans les articles suivants : M. Basseville et al., « Sequential detection of abrupt changes in spectral characteristics of digital signals » (IEEE Trans. on Information Theory, 1983, Vol. IT-29, n° 5, pages 708-723) ; R. Andre-Obrecht, « A new statistical approach for the automatic segmentation of continuous speech signals » (IEEE Trans. on Acous., Speech and Sig. Proc., Vol. 36, N°1, janvier 1988) ; et C. MURGIA et al., « An algorithm for the estimation of glottal closure instants using the sequential detection of abrupt changes in speech signals » (Signal Processing VII, 1994, pages 1685-1688).

La fréquence fondamentale estimée F_0 fait l'objet d'une quantification, par exemple scalaire, par un module 5, qui fournit au multiplexeur de sortie 6 un index iF de quantification de la fréquence fondamentale pour chaque trame du signal.

Le codeur utilise des modélisations paramétriques cepstrales pour représenter une enveloppe supérieure et une enveloppe inférieure du spectre du signal audio. La première étape de la transformation cepstrale consiste à appliquer au module du spectre du signal une fonction de compression

spectrale, qui peut être une fonction logarithmique ou en racine. Le module 8 du codeur opère ainsi, pour chaque valeur $X(i)$ du spectre du signal ($0 \leq i < N$), la transformation suivante :

$$LX(i) = \text{Log}(|X(i)|) \quad (4)$$

5 dans le cas d'une compression logarithmique ou

$$LX(i) = |X(i)|^\gamma \quad (5)$$

dans le cas d'une compression en racine, γ étant un exposant compris entre 0 et 1.

10 Le spectre comprimé LX du signal audio est traité par un module 9 qui extrait des amplitudes spectrales associées aux harmoniques du signal correspondant aux multiples de la fréquence fondamentale estimée F_0 . Ces amplitudes sont ensuite interpolées par un module 10 afin d'obtenir une enveloppe supérieure comprimée notée LX_{sup} .

15 Il est à noter que la compression spectrale pourrait de façon équivalente être effectuée après la détermination des amplitudes associées aux harmoniques. Elle pourrait également être effectuée après l'interpolation, ce qui ne ferait que modifier la forme des fonctions d'interpolation.

20 Le module 9 d'extraction des maxima tient compte de l'éventuelle variation de la fréquence fondamentale sur la trame d'analyse, des erreurs que peut commettre le détecteur 4, ainsi que des imprécisions liées au caractère discret de l'échantillonnage en fréquence. Pour cela, la recherche des amplitudes des pics spectraux ne consiste pas simplement à prendre les valeurs $LX(i)$ correspondant aux index i tels que $i.F_e/2N$ soit la fréquence la plus proche d'une harmonique de fréquence $k.F_0$ ($k \geq 1$). L'amplitude spectrale
25 retenue pour une harmonique d'ordre k est un maximum local du module du spectre au voisinage de la fréquence $k.F_0$ (cette amplitude est obtenue directement sous forme comprimée lorsque la compression spectrale 8 est effectuée avant l'extraction des maxima 9).

30 Les figures 4 et 5 montrent un exemple de forme du spectre comprimé LX , où on voit que les amplitudes maximales des pics harmoniques ne coïncident pas nécessairement avec les amplitudes correspondant aux multiples entiers de la fréquence fondamentale estimée F_0 . Les flancs des pics étant assez raides, une petite erreur de positionnement de la fréquence fondamentale F_0 , amplifiée par l'indice d'harmonique k , peut distordre

fortement l'enveloppe supérieure estimée du spectre et provoquer une mauvaise modélisation de la structure formantique du signal. Par exemple, prendre directement l'amplitude spectrale pour la fréquence $3.F_0$ dans le cas des figures 4 et 5 produirait une erreur importante dans l'extraction de l'enveloppe supérieure au voisinage de l'harmonique d'ordre $k = 3$, alors qu'il s'agit d'une zone énergétiquement importante dans l'exemple dessiné. En effectuant l'interpolation à partir du véritable maximum, on évite ce genre d'erreur d'estimation de l'enveloppe supérieure.

Dans l'exemple représenté sur la figure 4, l'interpolation est effectuée entre des points dont l'abscisse est la fréquence correspondant au maximum de l'amplitude d'un pic spectral, et dont l'ordonnée est ce maximum, avant ou après compression.

L'interpolation effectuée pour calculer l'enveloppe supérieure LX_{sup} est une simple interpolation linéaire. Bien entendu une autre forme d'interpolation pourrait être utilisée (par exemple polynomiale ou spline).

Dans la variante préférée représentée sur la figure 5, l'interpolation est effectuée entre des points dont l'abscisse est une fréquence $k.F_0$ multiple de la fréquence fondamentale (en fait la fréquence la plus proche dans le spectre discret) et dont l'ordonnée est l'amplitude maximale, avant ou après compression, du spectre au voisinage de cette fréquence multiple.

En comparant les figures 4 et 5, on peut voir que le mode d'extraction selon la figure 5, qui repositionne les pics sur les fréquences harmoniques, conduit à une meilleure précision sur l'amplitude des pics que le décodeur attribuera aux fréquences multiples de la fréquence fondamentale. Il peut se produire un léger déplacement en fréquence de la position de ces pics, ce qui n'est pas perceptuellement très important et n'est d'ailleurs pas évité non plus dans le cas de la figure 4. Dans le cas de la figure 4, les points d'ancrage pour l'interpolation sont confondus avec les sommets des pics harmoniques. Dans le cas de la figure 5, on impose que ces points d'ancrage se trouvent précisément aux fréquences multiples de la fréquence fondamentale, leurs amplitudes correspondant à celles des pics.

L'intervalle de recherche du maximum d'amplitude associé à une harmonique de rang k est centré sur l'index i de la fréquence de la TFR la plus proche de $k.F_0$, c'est-à-dire $i = \left\lfloor 2Nk \frac{F_0}{F_e} + \frac{1}{2} \right\rfloor$, où $\lfloor a \rfloor$ désigne l'entier égal ou

immédiatement inférieur au nombre a . La largeur de cet intervalle de recherche dépend de la fréquence d'échantillonnage F_e , de la taille $2N$ de la TFR et de la gamme de variation possible de la fréquence fondamentale. Cette largeur est typiquement de l'ordre d'une dizaine de fréquences avec les exemples de valeurs précédemment considérés. On peut la rendre réglable en fonction de la valeur F_0 de la fréquence fondamentale et du numéro k de l'harmonique.

Afin d'améliorer la résolution dans les basses fréquences et donc de représenter plus fidèlement les amplitudes des harmoniques dans cette zone, une distorsion non-linéaire de l'échelle des fréquences est opérée sur l'enveloppe supérieure comprimée par un module 12 avant que le module 13 effectue la transformée de Fourier rapide inverse (TFRI) fournissant les coefficients cepstraux cx_sup .

La distorsion non-linéaire permet de minimiser plus efficacement l'erreur de modélisation. Elle est par exemple effectuée selon une échelle de fréquences de type Mel ou Bark. Cette distorsion peut éventuellement dépendre de la fréquence fondamentale estimée F_0 . La figure 1 illustre le cas de l'échelle Mel. La relation entre les fréquences F du spectre linéaire, exprimées en hertz, et les fréquences F' de l'échelle Mel est la suivante :

$$F' = \frac{1000}{\log_{10}(2)} \times \log_{10} \left(1 + \frac{F}{1000} \right) \quad (6)$$

Afin de limiter le débit de transmission, une troncature des coefficients cepstraux cx_sup est effectuée. Le module de TFRI 13 a besoin de calculer seulement un vecteur cepstral de NCS coefficients cepstraux d'ordres 0 à NCS-1. A titre d'exemple, NCS peut être égal à 16.

Un post-filtrage dans le domaine cepstral, appelé post-liffrage, est appliqué par un module 15 à l'enveloppe supérieure comprimée LX_sup . Ce post-liffrage correspond à une manipulation des coefficients cepstraux cx_sup délivrés par le module de TRFI 13, qui correspond approximativement à un post-filtrage de la partie harmonique du signal par une fonction de transfert ayant la forme classique :

$$H(z) = \frac{(1 - \mu z^{-1}) A(z/\gamma_1)}{A(z/\gamma_2)} \quad (7)$$

où $A(z)$ est la fonction de transfert d'un filtre de prédiction linéaire du signal audio, γ_1 et γ_2 sont des coefficients compris entre 0 et 1, et μ est un coefficient de préaccentuation éventuellement nul. La relation entre le coefficient post-liffré

d'ordre i , noté $c_p(i)$, et le coefficient cepstral correspondant $c(i) = cx_sup(i)$ délivré par le module 13 est alors :

$$c_p(0) = c(0)$$

$$c_p(i) = \left(1 + \gamma_2^i - \gamma_1^i\right) c(i) - \frac{\mu^i}{i} \quad \text{pour } i > 0 \quad (8)$$

Le coefficient de préaccentuation optionnel μ peut être contrôlé en posant comme contrainte de préserver la valeur du coefficient cepstral $cx_sup(1)$ relatif à la pente. En effet, la valeur $c(1) = cx_sup(1)$ d'un bruit blanc filtré par le filtre de préaccentuation correspond au coefficient de préaccentuation. On peut ainsi choisir ce dernier de la façon suivante : $\mu = (\gamma_2 - \gamma_1) \cdot c(1)$.

10 Après le post-lifre 15, un module de normalisation 16 modifie encore les coefficients cepstraux en imposant la contrainte de modélisation exacte d'un point du spectre initial, qui est de préférence le point le plus énergétique parmi les maxima spectraux extraits par le module 9. En pratique, cette normalisation modifie seulement la valeur du coefficient $c_p(0)$.

15 Le module de normalisation 16 fonctionne de la façon suivante : il recalcule une valeur du spectre synthétisé à la fréquence du maximum indiqué par le module 9, par transformée de Fourier des coefficients cepstraux tronqués et post-lifrés, en tenant compte de la distorsion non-linéaire de l'axe des fréquences ; il détermine un gain de normalisation g_N par la différence
20 logarithmique entre la valeur du maximum fournie par le module 9 et cette valeur recalculée ; et il ajoute le gain g_N au coefficient cepstral post-lifré $c_p(0)$. Cette normalisation peut être vue comme faisant partie du post-lifrage.

Les coefficients cepstraux post-lifrés et normalisés font l'objet d'une quantification par un module 18 qui transmet des index de quantification
25 correspondants $icxs$ au multiplexeur de sortie 6 du codeur.

Le module 18 peut fonctionner par quantification vectorielle à partir de vecteurs cepstraux formés de coefficients post-lifrés et normalisés, notés ici $cx[n]$ pour la trame de signal de rang n . A titre d'exemple, le vecteur cepstral $cx[n]$ de $NCS = 16$ coefficients cepstraux $cx[n,0], cx[n,1], \dots, cx[n,NCS-1]$ est
30 distribué en quatre sous-vecteurs cepstraux contenant chacun quatre coefficients d'ordres consécutifs. Le vecteur cepstral $cx[n]$ peut être traité par les moyens représentés sur la figure 6, faisant partie du module de

quantification 18. Ces moyens mettent en œuvre, pour chaque composante $cx[n,i]$, un prédicteur de la forme :

$$cx_p[n,i] = (1 - \alpha(i)).rcx[n,i] + \alpha(i).rcx[n-1,i] \quad (9)$$

où $rcx[n]$ désigne un vecteur résiduel de prédiction pour la trame de rang n dont les composantes sont respectivement notées $rcx[n,0]$, $rcx[n,1]$, ..., $rcx[n,NCS-1]$, et $\alpha(i)$ désigne un coefficient de prédiction choisi pour être représentatif d'une corrélation inter-trame supposée. Après quantification des résidus, ce vecteur résiduel est défini par :

$$rcx[n,i] = \frac{cx[n,i] - \alpha(i).rcx_q[n-1,i]}{2 - \alpha(i)} \quad (10)$$

où $rcx_q[n-1]$ désigne le vecteur résiduel quantifié pour la trame de rang $n-1$, dont les composantes sont respectivement notées $rcx_q[n,0]$, $rcx_q[n,1]$, ..., $rcx_q[n,NCS-1]$.

Le numérateur de la relation (10) est obtenu par un soustracteur 20, dont les composantes du vecteur de sortie sont divisées par les quantités $2 - \alpha(i)$ en 21. Aux fins de la quantification, le vecteur résiduel $rcx[n]$ est subdivisé en quatre sous-vecteurs, correspondant à la subdivision en quatre sous-vecteurs cepstraux. Sur la base d'un dictionnaire obtenu par apprentissage préalable, l'unité 22 procède à la quantification vectorielle de chaque sous-vecteur du vecteur résiduel $rcx[n]$. Cette quantification peut consister, pour chaque sous-vecteur $srcx[n]$, à sélectionner dans le dictionnaire le sous-vecteur quantifié $srcx_q[n]$ qui minimise l'erreur quadratique $\|srcx[n] - srcx_q[n]\|^2$. L'ensemble $icxs$ des index de quantification icx , correspondant aux adresses dans le ou les dictionnaires des sous-vecteurs résiduels quantifiés $srcx_q[n]$, est fourni au multiplexeur de sortie 6.

L'unité 22 délivre également les valeurs des sous-vecteurs résiduels quantifiés, qui forment le vecteur $rcx_q[n]$. Celui-ci est retardé d'une trame en 23, et ses composantes sont multipliées par les coefficients $\alpha(i)$ en 24 pour fournir le vecteur à l'entrée négative du soustracteur 20. Ce dernier vecteur est d'autre part fourni à un additionneur 25, dont l'autre entrée reçoit un vecteur formé par les composantes du résidu quantifié $rcx_q[n]$ respectivement multipliées par les quantités $1 - \alpha(i)$ en 26. L'additionneur 25 délivre ainsi le vecteur cepstral quantifié $cx_q[n]$ que récupérera le décodeur.

Le coefficient de prédiction $\alpha(i)$ peut être optimisé séparément pour

chacun des coefficients cepstraux. Les dictionnaires de quantification peuvent aussi être optimisés séparément pour chacun quatre sous-vecteurs cepstraux. D'autre part, il est possible, de façon connue en soi, de normaliser les vecteurs cepstraux avant d'appliquer le schéma de prédiction/quantification, à partir de la variance des cepstres.

Il est à noter que le schéma ci-dessus de quantification des coefficients cepstraux peut n'être appliqué que pour certaines seulement des trames. Par exemple, on peut prévoir un second mode de quantification ainsi qu'un processus de sélection de celui des deux modes qui minimise un critère de moindres carrés avec les coefficients cepstraux à quantifier, et transmettre avec les index de quantification de la trame un bit indiquant lequel des deux modes a été sélectionné.

Les coefficients cepstraux quantifiés $cx_sup_q = cx_q[n]$ fournis par l'additionneur 25 sont adressés à un module 28 qui recalcule les amplitudes spectrales associées à une ou plusieurs des harmoniques de la fréquence fondamentale F_0 (figure 1). Ces amplitudes spectrales sont par exemple calculées sous forme comprimée, en appliquant la transformée de Fourier aux coefficients cepstraux quantifiés en tenant compte de la distorsion non-linéaire de l'échelle des fréquences utilisée dans la transformation cepstrale. Les amplitudes ainsi recalculées sont fournies à un module d'adaptation 29 qui les compare à des amplitudes de maxima déterminées par le module d'extraction 9.

Le module d'adaptation 29 contrôle le post-lifre 15 de façon à minimiser un écart de module entre le spectre du signal audio et les valeurs de module correspondantes calculées en 28. Cet écart de module peut être exprimé par une somme de valeurs absolues de différences d'amplitudes, comprimées ou non, correspondant à une ou plusieurs des fréquences harmoniques. Cette somme peut être pondérée en fonction des amplitudes spectrales associées à ces fréquences.

De façon optimale, l'écart de module pris en compte dans l'adaptation du post-lifrage tiendrait compte de toutes les harmoniques du spectre. Cependant, afin de réduire la complexité de l'optimisation, le module 28 peut ne resynthétiser les amplitudes spectrales que pour une ou plusieurs fréquences multiples de la fréquence fondamentale F_0 , sélectionnées sur la base de l'importance du module du spectre en valeur absolue. Le module

d'adaptation 29 peut par exemple considérer les trois pics spectraux les plus intenses dans le calcul de l'écart de module à minimiser.

Dans une autre réalisation, le module d'adaptation 29 estime une courbe de masquage spectral du signal audio au moyen d'un modèle psychoacoustique, et les fréquences prises en compte dans le calcul de l'écart de module à minimiser sont sélectionnées sur la base de l'importance du module du spectre relativement à la courbe de masquage (on peut par exemple prendre les trois fréquences pour lesquelles le module du spectre dépasse le plus de la courbe de masquage). Différentes méthodes classiques sont utilisables pour calculer la courbe de masquage à partir du signal audio. On peut par exemple utiliser celle développée par J.D. Johnston (« Transform Coding of Audio Signals Using Perceptual Noise Criteria », IEEE Journal on Selected Area in Communications, Vol. 6, No. 2, février 1988).

Pour réaliser l'adaptation du post-lifrage, le module 29 peut utiliser un modèle d'identification de filtre. Une méthode plus simple consiste à prédéfinir un ensemble de jeux de paramètres de post-lifrage, c'est-à-dire un ensemble de couples γ_1, γ_2 dans le cas d'un post-lifrage selon les relations (8), à effectuer les opérations incombant aux modules 15, 16, 18 et 28 pour chacun de ces jeux de paramètres, et à retenir celui des jeux de paramètres qui conduit à l'écart de module minimal entre le spectre du signal et les valeurs recalculées. Les index de quantification fournis par le module 18 sont alors ceux qui se rapportent au meilleur jeu de paramètres.

Par un processus analogue à celui de l'extraction des coefficients cx_sup représentant l'enveloppe supérieure comprimée LX_sup du spectre du signal, le codeur détermine des coefficients cx_inf représentant une enveloppe inférieure comprimée LX_inf . Un module 30 extrait du spectre comprimé LX des amplitudes spectrales associées à des fréquences situées dans des zones du spectre intermédiaires par rapport aux fréquences multiples de la fréquence fondamentale estimée F_0 .

Dans l'exemple illustré par les figures 4 et 5, chaque amplitude associée à une fréquence située dans une zone intermédiaire entre deux harmoniques successives $k.F_0$ et $(k+1).F_0$ correspond simplement au module du spectre pour la fréquence $(k+1/2).F_0$ située au milieu de l'intervalle séparant les deux harmoniques. Dans une autre réalisation, cette amplitude pourrait être une moyenne du module du spectre sur une petite plage entourant cette

fréquence $(k+1/2).F_0$.

Un module 31 procède à une interpolation, par exemple linéaire, des amplitudes spectrales associées aux fréquences situées dans les zones intermédiaires pour obtenir l'enveloppe inférieure comprimée LX_inf.

5 La transformation cepstrale appliquée à cette enveloppe inférieure comprimée LX_inf est effectuée suivant une échelle de fréquences résultant d'une distorsion non-linéaire appliquée par un module 32. Le module de TFRI 33 calcule un vecteur cepstral de NCI coefficients cepstraux cx_inf d'ordres 0 à NCI-1 représentant l'enveloppe inférieure. NCI est un nombre qui peut être
10 sensiblement plus petit que NCS, par exemple NCI = 4.

La transformation non-linéaire de l'échelle des fréquences pour la transformation cepstrale de l'enveloppe inférieure peut être réalisée vers une échelle plus fine aux hautes fréquences qu'aux basses fréquences, ce qui permet avantageusement de bien modéliser les composantes non-voisées du
15 signal aux hautes fréquences. Toutefois, pour assurer une homogénéité de représentation entre l'enveloppe supérieure et l'enveloppe inférieure, on pourra préférer adopter dans le module 32 la même échelle que dans le module 12 (Mel dans l'exemple considéré).

Les coefficients cepstraux cx_inf représentant l'enveloppe inférieure comprimée sont quantifiés par un module 34, qui peut fonctionner de la même
20 manière que le module 18 de quantification des coefficients cepstraux représentant l'enveloppe supérieure comprimée. Dans le cas considéré, où on se limite à NCI = 4 coefficients cepstraux pour l'enveloppe inférieure, le vecteur ainsi formé est soumis à une quantification vectorielle de résidu de prédiction,
25 effectuée par des moyens identiques à ceux représentés sur la figure 6 mais sans subdivision en sous-vecteurs. L'index de quantification icx = icxi déterminé par le quantificateur vectoriel 22 pour chaque trame relativement aux coefficients cx_inf est fourni au multiplexeur de sortie 6 du codeur.

Le codeur représenté sur la figure 1 ne comporte aucun dispositif
30 particulier pour coder les phases du spectre aux harmoniques du signal audio.

En revanche, il comporte des moyens 36-40 pour coder une information temporelle liée à la phase de la composante non-harmonique représentée par l'enveloppe inférieure.

Un module 36 de décompression spectrale et un module 37 de TFRI
35 forment une estimation temporelle de la trame de la composante non-

harmonique. Le module 36 applique une fonction de décompression réciproque de la fonction de compression appliquée par le module 8 (c'est-à-dire une exponentielle ou une fonction puissance $1/\gamma$) à l'enveloppe inférieure comprimée LX_inf produite par le module d'interpolation 31. Ceci fournit le
5 module de la trame estimée de la composante non-harmonique, dont la phase est prise égale à celle ϕ_X du spectre du signal X sur la trame. La transformée de Fourier inverse effectuée par le module 37 fournit la trame estimée de la composante non-harmonique.

Le module 38 subdivise cette trame estimée de la composante non-harmonique en plusieurs segments temporels. La trame délivrée par le module
10 37 se composant de $2N = 512$ échantillons pondérés comme illustré par la figure 3, le module 38 considère seulement les $N/2 = 128$ premiers échantillons et les $N/2 = 128$ derniers échantillons, et les subdivise par exemple en huit segments de 32 échantillons consécutifs représentant chacun 4 ms de
15 signal.

Pour chaque segment, le module 38 calcule l'énergie égale à la somme des carrés des échantillons, et forme un vecteur E1 formé de huit composantes réelles positives égales aux huit énergies calculées. La plus
20 grande de ces huit énergies, notée EM, est également déterminée pour être fournie, avec le vecteur E1, à un module de normalisation 39. Celui-ci divise chaque composante du vecteur E1 par EM, de sorte que le vecteur normalisé Emix est formé de huit composantes comprises entre 0 et 1. C'est ce vecteur normalisé Emix, ou vecteur de pondération, qui est soumis à la quantification par le module 40. Celui-ci peut opérer une quantification vectorielle avec un
25 dictionnaire déterminé lors d'un apprentissage préalable. L'index de quantification iEm est fourni par le module 40 au multiplexeur de sortie 6 du codeur.

La figure 7 montre une variante de réalisation des moyens employés par le codeur de la figure 1 pour déterminer le vecteur Emix de pondération
30 énergétique de la trame de la composante non-harmonique. Les modules 36, 37 de décompression spectrale et de TFRI fonctionnent comme ceux qui portent les mêmes références sur la figure 1. Un module de sélection 42 est ajouté pour déterminer la valeur du module du spectre soumis à la transformée de Fourier inverse 37. Sur la base de la fréquence fondamentale estimée F_0 , le
35 module 42 identifie des régions harmoniques et des régions non-harmoniques

du spectre du signal audio. Par exemple, une fréquence sera considérée comme appartenant à une région harmonique si elle se trouve dans un intervalle de fréquences centré sur une harmonique $k.F_0$ et de largeur correspondant à une largeur de raie spectrale synthétisée, et à une région non-harmonique sinon. Dans les régions non-harmoniques, le signal complexe soumis à la TFR 37 est égal à la valeur du spectre, c'est-à-dire que son module et sa phase correspondent aux valeurs $|X|$ et ϕ_X fournis par le module

de TFR 3. Dans les régions harmoniques, ce signal complexe a la même phase ϕ_X que le spectre et un module donné par l'enveloppe inférieure après décompression spectrale 36. Cette façon de procéder selon la figure 7 procure une modélisation plus précise des régions non-harmoniques.

Le décodeur représenté sur la figure 8 comprend un démultiplexeur d'entrée 45 qui extrait du flux binaire Φ , issu d'un codeur selon la figure 1, les index iF , $icxs$, $icxi$, iEm de quantification de la fréquence fondamentale F_0 , des coefficients cepstraux représentant l'enveloppe supérieure comprimée, des coefficients représentant l'enveloppe inférieure comprimée, et du vecteur de pondération $Emix$, et les distribue respectivement à des modules 46, 47, 48 et 49. Ces modules 46-49 comportent des dictionnaires de quantification semblables à ceux des modules 5, 18, 34 et 40 de la figure 1, afin de restituer les valeurs des paramètres quantifiés. Les modules 47 et 48 ont des dictionnaires pour former les résidus de prédiction quantifiés $rcx_q[n]$, et ils en déduisent les vecteurs cepstraux quantifiés $cx_q[n]$ avec des éléments identiques aux éléments 23-26 de la figure 6. Ces vecteurs cepstraux quantifiés $cx_q[n]$ fournissent les coefficients cepstraux cx_sup_q et cx_inf_q traités par le décodeur.

Un module 51 calcule la transformée de Fourier rapide des coefficients cepstraux cx_sup pour chaque trame de signal. L'échelle des fréquences du spectre comprimé qui en résulte est modifiée non-linéairement par un module 52 appliquant la transformation non-linéaire réciproque de celle du module 12 de la figure 1, et qui fournit l'estimation LX_sup de l'enveloppe supérieure comprimée. Une décompression spectrale de LX_sup , opérée par un module 53, fournit l'enveloppe supérieure X_sup comportant les valeurs estimées du module du spectre aux fréquences multiples de la fréquence fondamentale F_0 . Le module 54 synthétise l'estimation spectrale X_v de la composante

harmonique du signal audio, par une somme de raies spectrales centrées sur les fréquences multiples de la fréquence fondamentale F_0 et dont les amplitudes (en module) sont celles données par l'enveloppe supérieure X_{sup} .

Bien que le flux numérique d'entrée Φ ne comporte pas d'informations
5 spécifiques sur la phase du spectre du signal aux harmoniques de la fréquence fondamentale, le décodeur de la figure 8 est capable d'extraire de l'information sur cette phase à partir des coefficients cepstraux $cx_{sup,q}$ représentant l'enveloppe supérieure comprimée. Cette information de phase est utilisée pour affecter une phase $\phi(k)$ à chacune des raies spectrales déterminées par le
10 module 54 dans l'estimation de la composante harmonique du signal.

En première approximation, le signal de parole peut être considéré comme étant à phase minimale. D'autre part, il est connu que l'information de phase minimale peut se déduire facilement d'une modélisation cepstrale. Cette information de phase minimale est donc calculée pour chaque fréquence
15 harmonique. L'hypothèse de phase minimale signifie que l'énergie du signal synthétisé est localisée au début de chaque période de la fréquence fondamentale F_0 .

Pour être plus proche d'un signal de parole réel, on introduit un peu de dispersion au moyen d'un post-lifrage spécifique des cepstres lors de la
20 synthèse de la phase. Avec ce post-lifrage, effectué par le module 55 de la figure 8, il est possible d'accentuer les résonances formantiques de l'enveloppe et donc de contrôler la dispersion des phases. Ce post-lifrage est par exemple de la forme (8).

Pour limiter les ruptures de phase, il est préférable de lisser les
25 coefficients cepstraux post-liftrés, ce qui est effectué par le module 56. Le module 57 déduit des coefficients cepstraux post-liftrés et lissés la phase minimale affectée à chaque raie spectrale représentant un pic harmonique du spectre.

Les opérations effectuées par les modules 56, 57 de lissage et
30 d'extraction de la phase minimale sont illustrées par l'organigramme de la figure 9. Le module 56 examine les variations des coefficients cepstraux pour appliquer un lissage moins important en présence de variations brusques qu'en présence de variations lentes. Pour cela, il effectue le lissage des coefficients cepstraux au moyen d'un facteur d'oubli λ_c choisi en fonction d'une
35 comparaison entre un seuil d_{th} et une distance d entre deux jeux successifs de

coefficients cepstraux post-liftrés. Le seuil d_{th} est lui-même adapté en fonction des variations des coefficients cepstraux.

La première étape 60 consiste à calculer la distance d entre les deux vecteurs successifs relatifs aux trames $n-1$ et n . Ces vecteurs, notés ici
5 $cxp[n-1]$ et $cxp[n]$, correspondent pour chaque trame à l'ensemble des NCS coefficients cepstraux post-liftrés représentant l'enveloppe supérieure comprimée. La distance utilisée peut notamment être la distance euclidienne entre les deux vecteurs ou encore une distance quadratique.

Deux lissages sont d'abord effectués, respectivement au moyen de
10 facteurs d'oubli λ_{min} et λ_{max} , pour déterminer une distance minimale d_{min} et une distance maximale d_{max} . Le seuil d_{th} est ensuite déterminé à l'étape 70 comme étant situé entre les distances minimale et maximale d_{min} , d_{max} :
$$d_{th} = \beta \cdot d_{max} + (1-\beta) \cdot d_{min}$$
, le coefficient β étant par exemple égal à 0,5.

Dans l'exemple représenté, les facteurs d'oubli λ_{min} et λ_{max} sont eux-
15 mêmes sélectionnés parmi deux valeurs distinctes, respectivement λ_{min1} , λ_{min2} et λ_{max1} , λ_{max2} comprises entre 0 et 1, les indices λ_{min1} , λ_{max1} étant chacun sensiblement plus près de 0 que les indices λ_{min2} , λ_{max2} . Si $d > d_{min}$ (test 61), le facteur d'oubli λ_{min} est égal à λ_{min1} (étape 62) ; sinon il est pris égal à λ_{min2} (étape 63). A l'étape 64, la distance minimale d_{min} est prise égale à
20 $\lambda_{min} \cdot d_{min} + (1-\lambda_{min}) \cdot d$. Si $d > d_{max}$ (test 65), le facteur d'oubli λ_{max} est égal à λ_{max1} (étape 66) ; sinon il est pris égal à λ_{max2} (étape 67). A l'étape 68, la distance minimale d_{max} est prise égale à $\lambda_{max} \cdot d_{max} + (1-\lambda_{max}) \cdot d$.

Si la distance d entre les deux vecteurs cepstraux consécutifs est plus grande que le seuil d_{th} (test 71), on adopte pour le facteur d'oubli λ_c une valeur
25 λ_{c1} relativement proche de 0 (étape 72). On considère dans ce cas que le signal correspondant est de type non stationnaire, de sorte qu'il n'y a pas lieu de conserver une grande mémoire des coefficients cepstraux antérieurs. Si $d \leq d_{th}$, on adopte à l'étape 73 pour le facteur d'oubli λ_c une valeur λ_{c2} moins proche de 0 afin de lisser davantage les coefficients cepstraux. Le lissage est
30 effectué à l'étape 74, où le vecteur $cxl[n]$ de coefficients lissés pour la trame courante n est déterminé par :

$$cxl[n] = \lambda_c \cdot cxl[n-1] + (1 - \lambda_c) \cdot cyp[n] \quad (11)$$

Le module 57 calcule ensuite les phases minimales $\varphi(k)$ associées aux harmoniques $k.F_0$. De façon connue, la phase minimale pour une harmonique d'ordre k est donnée par :

$$\varphi(k) = -2 \cdot \sum_{m=1}^{NCS-1} cxl[n,m] \cdot \sin(2\pi mk F_0 / F_e) \quad (12)$$

où $cxl[n,m]$ désigne le coefficient cepstral lissé d'ordre m pour la trame n .

A l'étape 75, l'index d'harmonique k est initialisé à 1. Pour initialiser le calcul de la phase minimale affectée à l'harmonique k , la phase $\varphi(k)$ et l'index cepstral m sont initialisés respectivement à 0 et 1 à l'étape 76. A l'étape 77, le module 57 ajoute à la phase $\varphi(k)$ la quantité $-2 \cdot cxl[n,m] \cdot \sin(2\pi mk F_0 / F_e)$. L'index cepstral m est incrémenté à l'étape 78 et comparé à NCS à l'étape 79. Les étapes 77 et 78 sont répétées tant que $m < NCS$. Quand $m = NCS$, le calcul de la phase minimale est terminé pour l'harmonique k , et l'index k est incrémenté à l'étape 80. Le calcul de phases minimales 76-79 est renouvelé pour l'harmonique suivante tant que $k.F_0 < F_e/2$ (test 81).

Dans l'exemple de réalisation selon la figure 8, le module 54 tient compte d'une phase constante sur la largeur de chaque raie spectrale, égale à la phase minimale $\varphi(k)$ fournie pour l'harmonique correspondante k par le module 57.

L'estimation X_v de la composante harmonique est synthétisée par sommation de raies spectrales positionnées aux fréquences harmoniques de la fréquence fondamentale F_0 . Lors de cette synthèse, on peut positionner les raies spectrales sur l'axe des fréquences avec une résolution supérieure à la résolution de la transformée de Fourier. Pour cela, on précalcule une fois pour toutes une raie spectrale de référence selon la résolution supérieure. Ce calcul peut consister en une transformée de Fourier de la fenêtre d'analyse f_A avec une taille de transformée de 16384 points, procurant une résolution de 0,5 Hz par point. La synthèse de chaque raie harmonique est alors effectuée par le module 54 en positionnant sur l'axe des fréquences la raie de référence à haute résolution, et en sous-échantillonnant cette raie spectrale de référence pour se ramener à la résolution de 16,625 Hz de la transformée de Fourier sur 512 points. Ceci permet de positionner avec précision la raie spectrale.

Pour la détermination de l'enveloppe inférieure, le module de TFR 85 du décodeur de la figure 8 reçoit les NCI coefficients cepstraux quantifiés cx_inf_q d'ordres 0 à $NCI - 1$, et il les complète avantageusement par les $NCS - NCI$ coefficients cepstraux cx_sup_q d'ordre NCI à $NCS - 1$ représentant l'enveloppe supérieure. En effet, on peut estimer en première approximation que les variations rapides de l'enveloppe inférieure comprimée sont bien reproduites par celles de l'enveloppe supérieure comprimée. Dans une autre réalisation, le module de TFR 85 pourrait ne considérer que les NCI paramètres cepstraux cx_inf_q .

Le module 86 convertit l'échelle de fréquences de manière réciproque de la conversion opérée par le module 32 du codeur, afin de restituer l'estimation LX_inf de l'enveloppe inférieure comprimée, soumise au module de décompression spectrale 87. En sortie du module 87, le décodeur dispose d'une enveloppe inférieure X_inf comportant les valeurs du module du spectre dans les vallées situées entre les pics harmoniques.

Cette enveloppe X_inf va moduler le spectre d'une trame de bruit dont la phase est traitée en fonction du vecteur de pondération quantifié $Emix$ extrait par le module 49. Un générateur 88 délivre une trame de bruit normalisé dont les segments de 4 ms sont pondérés dans un module 89 conformément aux composantes normalisées du vecteur $Emix$ fourni par le module 49 pour la trame courante. Ce bruit est un bruit blanc filtré passe-haut pour tenir compte du faible niveau qu'a en principe la composante non-voisée aux basses fréquences. A partir du bruit pondéré en énergie, le module 90 forme des trames de $2N = 512$ échantillons en appliquant la fenêtre d'analyse f_A , l'insertion de 256 échantillons à zéro et la permutation circulaire pour la compensation de phase conformément à ce qui a été expliqué en référence à la figure 3. La transformée de Fourier de la trame résultante est calculée par le module TFR 91.

L'estimation spectrale X_{uv} de la composante non-harmonique est déterminée par le module de synthèse spectrale 92 qui effectue une pondération fréquence par fréquence. Cette pondération consiste à multiplier chaque valeur spectrale complexe fournie par le module de TFR 91 par la valeur de l'enveloppe inférieure X_inf obtenue pour la même fréquence par le module de décompression spectrale 87.

Les estimations spectrales X_v , X_{uv} des composantes harmonique

(voisée dans le cas d'un signal de parole) et non-harmonique (ou non-voisée) sont combinées par un module de mixage 95 contrôlé par un module 96 d'analyse du degré d'harmonicité (ou de voisement) du signal.

L'organisation de ces modules 95, 96 est illustrée par la figure 10. Le module d'analyse 96 comporte une unité 97 d'estimation d'un degré de voisement W dépendant de la fréquence, à partir duquel sont calculés quatre gains dépendant de la fréquence, à savoir deux gains g_v , g_{uv} contrôlant l'importance relative des composantes harmonique et non-harmonique dans le signal synthétisé, et deux gains g_{v_φ} , g_{uv_φ} utilisés pour bruiteur la phase de la composante harmonique.

Le degré de voisement $W(i)$ est une valeur à variation continue comprise entre 0 et 1 déterminée pour chaque index de fréquence i ($0 \leq i < N$) en fonction de l'enveloppe supérieure $X_{sup}(i)$ et de l'enveloppe inférieure $X_{inf}(i)$ obtenues pour cette fréquence i par les modules de décompression 53, 87. Le degré de voisement $W(i)$ est estimé par l'unité 97 pour chaque index de fréquence i correspondant à une harmonique de la fréquence fondamentale F_0 , à savoir $i = \left\lfloor 2Nk \frac{F_0}{F_e} + \frac{1}{2} \right\rfloor$ pour $k = 1, 2, \dots$, par une fonction croissante du rapport entre l'enveloppe supérieure X_{sup} et l'enveloppe inférieure X_{inf} à cette fréquence, par exemple selon la formule :

$$W(i) = \min \left\{ 1, \frac{10 \cdot \log_{10} [X_{sup}(i)/X_{inf}(i)]}{V_{th}(F_0)} \right\} \quad (13)$$

Le seuil $V_{th}(F_0)$ correspond à la dynamique moyenne calculée sur un spectre synthétique purement voisé à la fréquence fondamentale. Il est avantageusement choisi dépendant de la fréquence fondamentale F_0 .

Le degré de voisement $W(i)$ pour une fréquence autre que les fréquences harmoniques est obtenu simplement comme étant égal à celui estimé pour l'harmonique la plus proche.

Le gain $g_v(i)$, qui dépend de la fréquence, est obtenu en appliquant une fonction non-linéaire au degré de voisement $W(i)$ (bloc 98). Cette fonction non-linéaire a par exemple la forme représentée sur la figure 11 :

$$g_v(i) = 0 \quad \text{si } 0 \leq W(i) \leq W1$$
$$g_v(i) = \frac{W(i) - W1}{W2 - W1} \quad \text{si } W1 < W(i) < W2 \quad (14)$$

$$g_v(i) = 1 \quad \text{si } W2 \leq W(i) \leq 1$$

les seuils $W1$, $W2$ étant tels que $0 < W1 < W2 < 1$. Le gain g_{uv} peut être calculé de manière semblable au gain g_v (la somme des deux gains g_v , g_{uv} étant constante, par exemple égale à 1), ou déduit simplement de celui-ci par la relation $g_{uv}(i) = 1 - g_v(i)$, comme schématisé par le soustracteur 99 sur la figure 10.

Il est intéressant de pouvoir bruite la phase de la composante harmonique du signal à une fréquence donnée si l'analyse du degré de voisement montre que le signal est plutôt de type non-harmonique à cette fréquence. Pour cela, la phase ϕ_v de la composante harmonique mixée est le résultat d'une combinaison linéaire des phases ϕ_v , ϕ_{uv} des composantes harmonique et non-harmonique X_v , X_{uv} synthétisées par les modules 54, 92. Les gains g_{v_ϕ} , g_{uv_ϕ} respectivement appliqués à ces phases sont calculés à partir du degré de voisement W et pondérés également en fonction de l'index de fréquence i , étant donné que le bruitage de la phase n'est véritablement utile qu'au-delà d'une certaine fréquence.

Un premier gain g_{v1_ϕ} est calculé en appliquant une fonction non-linéaire au degré de voisement $W(i)$, comme schématisé par le bloc 100 sur la figure 10. Cette fonction non-linéaire peut avoir la forme représentée sur la figure 12 :

$$\begin{aligned} g_{v1_\phi}(i) &= G1 \quad \text{si } 0 \leq W(i) \leq W3 \\ g_{v1_\phi}(i) &= G1 + (1 - G1) \frac{W(i) - W3}{W4 - W3} \quad \text{si } W3 < W(i) < W4 \\ g_{v1_\phi}(i) &= 1 \quad \text{si } W4 \leq W(i) \leq 1 \end{aligned} \quad (15)$$

les seuils $W3$ et $W4$ étant tels que $0 < W3 < W4 < 1$, et le gain minimal $G1$ étant compris entre 0 et 1.

Un multiplieur 101 multiplie pour chaque fréquence d'index i le gain g_{v1_ϕ} par un autre gain g_{v2_ϕ} dépendant seulement de l'index de fréquence i , pour former le gain $g_{v_\phi}(i)$. Le gain $g_{v2_\phi}(i)$ dépend non-linéairement de l'index de fréquence i , par exemple comme indiqué sur la figure 13 :

$$\begin{aligned} g_{v2_\phi}(i) &= 1 \quad \text{si } 0 \leq i \leq i1 \\ g_{v2_\phi}(i) &= 1 - (1 - G2) \frac{i - i1}{i2 - i1} \quad \text{si } i1 < i < i2 \end{aligned} \quad (16)$$

$$g_{v2_φ}(i) = G2 \quad \text{si } i2 \leq i \leq 1$$

les index $i1$ et $i2$ étant tels que $0 < i1 < i2 \leq N$, et le gain minimal $G2$ étant compris entre 0 et 1. Le gain $g_{uv_φ}(i)$ peut être calculé simplement comme étant égal à $1 - g_{v_φ}(i) = 1 - g_{v1_φ}(i) \cdot g_{v2_φ}(i)$ (soustracteur 102 de la figure 10).

Le spectre complexe Y du signal synthétisé est produit par le module de mixage 95, qui réalise la relation de mixage suivante, pour $0 \leq i < N$:

$$Y(i) = g_v(i) \cdot |X_v(i)| \cdot \exp[j \phi_v(i)] + g_{uv}(i) \cdot X_{uv}(i) \quad (17)$$

$$\text{avec} \quad \phi_v(i) = g_{v_φ}(i) \cdot \phi_v(i) + g_{uv_φ}(i) \cdot \phi_{uv}(i) \quad (18)$$

où $\phi_v(i)$ désigne l'argument du nombre complexe $X_v(i)$ fourni par le module 54 pour la fréquence d'index i (bloc 104 de la figure 10), et $\phi_{uv}(i)$ désigne l'argument du nombre complexe $X_{uv}(i)$ fourni par le module 92 (bloc 105 de la figure 10). Cette combinaison est réalisée par les multiplieurs 106-110 et les additionneurs 111-112 représentés sur la figure 10.

Le spectre mixé $Y(i)$ pour $0 \leq i < 2N$ (avec $Y(2N-1-i) = Y(i)$) est ensuite transformé dans le domaine temporel par le module de TFRI 115 (figure 8). On ne retient que les $N/2 = 128$ premiers et les $N/2 = 128$ derniers échantillons de la trame de $2N = 512$ échantillons produite par le module 115, et on applique la permutation circulaire inverse de celle illustrée par la figure 3 pour obtenir la trame synthétisée de $N = 256$ échantillons pondérés par la fenêtre d'analyse f_A .

Les trames successivement obtenues de cette manière sont finalement traitées par le module de synthèse temporelle 116 qui forme le signal audio décodé \hat{x} .

Le module de synthèse temporelle 116 effectue une somme à recouvrement de trames modifiées par rapport à celles successivement évaluées en sortie du module 115. La modification peut être vue en deux étapes illustrées respectivement par les figures 14 et 15.

La première étape (figure 14) consiste à multiplier chaque trame 2' délivrée par le module de TFRI 115 par une fenêtre $1/f_A$ inverse de la fenêtre d'analyse f_A employée par le module 1 du codeur. Les échantillons de la trame 2" qui en résultent sont donc pondérés uniformément.

La seconde étape (figure 15) consiste à multiplier les échantillons de

cette trame 2" par une fenêtre de synthèse f_S vérifiant les propriétés suivantes :

$$f_S(N-L+i) + f_S(i) = A \quad \text{pour } 0 \leq i < L \quad (19)$$

$$f_S(i) = A \quad \text{pour } L \leq i < N-L \quad (20)$$

où A désigne une constante positive arbitraire, par exemple $A = 1$. La fenêtre de synthèse $f_S(i)$ croît progressivement de 0 à A pour i allant de 0 à L. C'est par exemple une demi-sinusoïde surélevée :

$$f_S(i) = \frac{A}{2} \cdot (1 - \cos[(i + 1/2)\pi/L]) \quad \text{pour } 0 \leq i < L \quad (21)$$

Après avoir repondéré chaque trame 2" par la fenêtre de synthèse f_S , le module 116 positionne les trames successives avec leurs décalages temporels de $M = 160$ échantillons et leurs recouvrements temporels de $L = 96$ échantillons, puis il effectue la somme des trames ainsi positionnées dans le temps. Du fait des propriétés (19) et (20) de la fenêtre de synthèse f_S , chaque échantillon du signal audio décodé \hat{x} ainsi obtenu est affecté d'un poids global uniforme, égal à A. Ce poids global provient de la contribution d'une trame unique si l'échantillon a dans cette trame un rang i tel que $L \leq i < N - L$, et comporte les contributions sommées de deux trames successives si $0 \leq i < L$ ou $N - L \leq i < N$.

On peut ainsi effectuer la synthèse temporelle de façon simple même si, comme dans le cas considéré, le recouvrement L entre deux trames successives est plus petit que la moitié de la taille N de ces trames.

Les deux étapes exposées ci-dessus pour la modification des trames de signal peuvent être fusionnées en une seule étape. Il suffit de précalculer une fenêtre composée $f_C(i) = f_S(i)/f_A(i)$, et de multiplier simplement les trames 2' de $N = 256$ échantillons délivrées par le module 115 par la fenêtre composée f_C avant d'effectuer la sommation à recouvrement.

La figure 16 montre l'allure de la fenêtre composée f_C dans le cas où la fenêtre d'analyse f_A est une fenêtre de Hamming et la fenêtre de synthèse f_S a la forme donnée par les relations (19) à (21).

D'autres formes de la fenêtre de synthèse f_S vérifiant les relations (19) et (20) peuvent être employées. Dans la variante de la figure 17, c'est une fonction affine par morceaux définie par :

$$f_S(i) = A \cdot i/L \quad \text{pour } 0 \leq i < L \quad (22)$$

Afin d'améliorer la qualité de codage du signal audio, le codeur de la figure 1 peut augmenter la cadence de formation et d'analyse des trames, afin de transmettre davantage de paramètres de quantification au décodeur. Dans la structure de trame représentée sur la figure 2, une trame de $N = 256$ échantillons (32 ms) est formée toutes les 20 ms. Ces trames de 256 échantillons pourraient être formées à une cadence supérieure, par exemple de 10 ms, deux trames successives ayant alors un décalage de $M/2 = 80$ échantillons et un recouvrement de 176 échantillons.

Dans ces conditions, on peut transmettre les jeux complets de paramètres de quantification iF , $icxs$, $icxi$, iEm pour seulement un sous-ensemble des trames, et transmettre pour les autres trames des paramètres permettant d'effectuer une interpolation adéquate au niveau du décodeur. Dans l'exemple envisagé ci-dessus, le sous-ensemble pour lequel des jeux de paramètres complets sont transmis peut être constitué par les trames de rang entier n , dont la périodicité est de $M/F_e = 20$ ms, et les trames pour lesquelles une interpolation est effectuée peuvent être celles de rang demi-entier $n + 1/2$ qui sont décalées de 10 ms par rapport aux trames du sous-ensemble.

Dans la réalisation illustrée par la figure 18, les notations $cx_q[n-1]$ et $cx_q[n]$ désignent des vecteurs cepstraux quantifiés déterminés, pour deux trames successives de rang entier, par le module de quantification 18 et/ou par le module de quantification 34. Ces vecteurs comprennent par exemple quatre coefficients cepstraux consécutifs chacun. Ils pourraient également comprendre davantage de coefficients cepstraux.

Un module 120 effectue une interpolation de ces deux vecteurs cepstraux $cx_q[n-1]$ et $cx_q[n]$, afin d'estimer une valeur intermédiaire $cx_i[n-1/2]$. L'interpolation effectuée par le module 120 peut être une simple moyenne arithmétique des vecteurs $cx_q[n-1]$ et $cx_q[n]$. En variante, le module 120 pourrait appliquer une formule d'interpolation plus sophistiquée, par exemple polynomiale, en se fondant également sur les vecteurs cepstraux obtenus pour des trames antérieures à la trame $n-1$. D'autre part, si plus d'une trame interpolée est intercalée entre deux trames consécutives de rang entier, l'interpolation tient compte de la position relative de chaque trame interpolée.

A l'aide des moyens précédemment décrits, le codeur calcule également les coefficients cepstraux $cx[n-1/2]$ relatifs à la trame de rang demi-entier. Dans le cas de l'enveloppe supérieure, ces coefficients cepstraux sont

ceux fournis par le module de TFRI 13 après post-liffrage 15 (par exemple avec les mêmes coefficients de post-liffrage que pour la trame précédente $n-1$) et normalisation 16. Dans le cas de l'enveloppe inférieure, les coefficients cepstraux $cx[n-1/2]$ sont ceux délivrés par le module de TFRI 33.

5 Un soustracteur 121 forme la différence $ecx[n-1/2]$ entre les coefficients cepstraux $cx[n-1/2]$ calculés pour la trame de rang demi-entier et les coefficients $cx_i[n-1/2]$ estimés par interpolation. Cette différence est fournie à un module de quantification 122 qui adresse des index de quantification $icx[n-1/2]$ au multiplexeur de sortie 6 du codeur. Le module 122
10 fonctionne par exemple par quantification vectorielle des erreurs d'interpolation $ecx[n-1/2]$ successivement déterminées pour les trames de rang demi-entier.

Cette quantification de l'erreur d'interpolation peut être effectuée par le codeur pour chacun des NCS + NCI coefficients cepstraux utilisés par le décodeur, ou seulement pour certains d'entre eux, typiquement ceux d'ordres
15 les plus petits.

Les moyens correspondants du décodeur sont illustrés par la figure 19. Le décodeur fonctionne essentiellement comme celui décrit en référence à la figure 8 pour déterminer les trames de signal de rang entier. Un module d'interpolation 124 identique au module 120 du codeur estime les coefficients
20 intermédiaires $cx_i[n-1/2]$ à partir des coefficients quantifiés $cx_q[n-1]$ et $cx_q[n]$ fournis par le module 47 et/ou le module 48 à partir des index $icxs$, $icxi$ extraits du flux Φ . Un module d'extraction de paramètres 125 reçoit l'index de quantification $icx[n-1/2]$ depuis le démultiplexeur d'entrée 45 du décodeur, et en déduit l'erreur d'interpolation quantifiée $ecx_q[n-1/2]$ à partir du même
25 dictionnaire de quantification que celui utilisé par le module 122 du codeur. Un additionneur 126 fait la somme des vecteurs cepstraux $cx_i[n-1/2]$ et $ecx_q[n-1/2]$ afin de fournir les coefficients cepstraux $cx[n-1/2]$ qui seront utilisés par le décodeur (modules 51-57, 95, 96, 115 et/ou modules 85-87, 92, 95, 96, 115) pour former la trame interpolée de rang $n-1/2$.

30 Si certains seulement des coefficients cepstraux ont fait l'objet d'une quantification d'erreur d'interpolation, les autres sont déterminés par le décodeur par une interpolation simple, sans correction.

Le décodeur peut également interpoler les autres paramètres F_0 , $Emix$ utilisés pour synthétiser les trames de signal. La fréquence fondamentale F_0
35 peut être interpolée linéairement, soit dans le domaine temporel, soit (de

préférence) directement dans le domaine fréquentiel. Pour l'interpolation éventuelle du vecteur de pondération énergétique E_{mix} , il convient d'effectuer l'interpolation après dénormalisation et en tenant compte bien entendu des décalages temporels entre trames.

5 Il est à noter qu'il est particulièrement avantageux, pour interpoler la représentation des enveloppes spectrales, d'effectuer cette interpolation dans le domaine cepstral. Contrairement à une interpolation effectuée sur d'autres paramètres, tels que les coefficients LSP (« Line Spectrum Pairs »), l'interpolation linéaire des coefficients cepstraux correspond à l'interpolation
10 linéaire des amplitudes spectrales comprimées.

Dans la variante représentée sur la figure 20, le codeur utilise les vecteurs cepstraux $cx_q[n]$, $cx_q[n-1]$, ..., $cx_q[n-r]$ et $cx_q[n-1/2]$ calculés pour les dernières trames passées ($r \geq 1$) pour identifier un filtre interpolateur optimal qui, lorsqu'on lui soumet les vecteurs cepstraux quantifiés $cx_q[n-r]$,
15 ..., $cx_q[n]$ relatifs aux trames de rang entier, délivre un vecteur cepstral interpolé $cx_i[n-1/2]$ qui présente une distance minimale avec le vecteur $cx[n-1/2]$ calculé pour la dernière trame de rang demi-entier.

Dans l'exemple représenté sur la figure 20, ce filtre interpolateur 128 est présent dans le codeur, et un soustracteur 129 retranche sa sortie
20 $cx_i[n-1/2]$ du vecteur cepstral calculé $cx[n-1/2]$. Un module de minimisation 130 détermine le jeu de paramètres $\{P\}$ du filtre interpolateur 128, pour lequel l'erreur d'interpolation $ecx[n-1/2]$ délivrée par le soustracteur 129 présente une norme minimale. Ce jeu de paramètres $\{P\}$ est adressé à un module de quantification 131 qui fournit un index de quantification correspondant iP au
25 multiplexeur de sortie 6 du codeur.

En fonction du débit alloué dans le flux Φ aux index de quantification des paramètres $\{P\}$ définissant le filtre interpolateur optimal 128, on pourra adopter une quantification plus ou moins fine de ces paramètres, ou une forme plus ou moins élaborée du filtre interpolateur, ou encore prévoir plusieurs filtres
30 interpolateurs quantifiés de manière distincte pour différents vecteurs de coefficients cepstraux.

Dans une réalisation simple, le filtre interpolateur 128 est linéaire, avec
 $r = 1$:

$$cx_i[n-1/2] = p.cx_q[n-1] + (1-p).cx_q[n] \quad (23)$$

35 et le jeu de paramètres $\{P\}$ se limite au coefficient p compris entre 0 et 1.

A partir des index iP de quantification des paramètres $\{P\}$ obtenus dans le flux binaire ϕ , le décodeur reconstruit le filtre interpolateur 128 (aux erreurs de quantification près), et traite les vecteurs spectraux $cx_q[n-r]$, ..., $cx_q[n]$ afin d'estimer les coefficients cepstraux $cx[n-1/2]$ utilisés pour synthétiser les trames de rang demi-entier.

De façon générale, le décodeur peut utiliser une méthode d'interpolation simple (sans transmission de paramètres de la part du codeur pour les trames de rang demi-entier), une méthode d'interpolation avec prise en compte d'une erreur d'interpolation quantifiée (selon les figures 17 et 18), ou une méthode d'interpolation avec un filtre interpolateur optimal (selon la figure 19) pour évaluer les trames de rang demi-entier en plus des trames de rang entier évaluées directement comme expliqué en référence aux figures 8 à 13. Le module 116 de synthèse temporelle peut alors combiner l'ensemble de ces trames évaluées pour former le signal synthétisé \hat{x} de la manière expliquée ci-après en référence aux figures 14, 21 et 22.

Comme dans la méthode de synthèse temporelle précédemment décrite, le module 116 effectue une somme à recouvrement de trames modifiées par rapport à celles successivement évaluées en sortie du module 115, et cette modification peut être vue en deux étapes dont la première est identique à celle précédemment décrite en référence à la figure 14 (diviser les échantillons de la trame 2' par la fenêtre d'analyse f_A).

La seconde étape (figure 21) consiste à multiplier les échantillons de la trame renormalisée 2'' par une fenêtre de synthèse f'_S vérifiant les propriétés suivantes :

$$f'_S(i) = 0 \quad \text{pour } 0 \leq i < N/2 - M/p \text{ et } N/2 + M/p \leq i < N \quad (24)$$

$$f'_S(i) + f'_S(i + M/p) = A \quad \text{pour } N/2 - M/p \leq i < N/2 \quad (25)$$

où A désigne une constante positive arbitraire, par exemple $A = 1$, et p est l'entier tel que le décalage temporel entre les trames successives (calculées directement et interpolées) soit de M/p échantillons, soit $p = 2$ dans l'exemple décrit. La fenêtre de synthèse $f'_S(i)$ croît progressivement pour i allant de $N/2 - M/p$ à $N/2$. C'est par exemple une sinusoïde surélevée sur l'intervalle $N/2 - M/p \leq i < N/2 + M/p$. En particulier, la fenêtre de synthèse f'_S peut être, sur cet intervalle, une fenêtre de Hamming (comme représenté sur la figure 21)

ou une fenêtre de Hanning.

La figure 21 montre les trames successives 2" repositionnées dans le temps par le module 116. Les hachures indiquent les portions éliminées des trames (fenêtre de synthèse à 0). On voit qu'en effectuant la somme à recouvrement des échantillons des trames successives, la propriété (25) assure une pondération homogène des échantillons du signal synthétisé.

Comme dans la méthode de synthèse illustrée par les figures 14 et 15, la procédure de pondération des trames obtenues par transformée de Fourier inverse des spectres Y peut être effectuée en une seule étape, avec une fenêtre composée $f'_C(i) = f'_S(i)/f_A(i)$. La figure 22 montre la forme de la fenêtre composée f'_C dans le cas où les fenêtres f_A et f'_S sont de type Hamming.

Comme la méthode de synthèse temporelle illustrée par les figures 14 à 17, celle illustrée par les figures 14, 21 et 22 permet de prendre en compte un recouvrement L entre deux trames d'analyse (pour lesquelles l'analyse est effectuée de façon complète) plus petit que la moitié que la taille N de ces trames. De façon générale, cette dernière méthode est applicable lorsque les trames d'analyse successives présentent des décalages temporels mutuels M de plus de N/2 échantillons (même éventuellement de plus de N échantillons si un très bas débit est requis), l'interpolation conduisant à un ensemble de trames dont les décalages temporels mutuels sont de moins de N/2 échantillons.

Les trames interpolées peuvent faire l'objet d'une transmission réduite de paramètres de codage, comme décrit précédemment, mais cela n'est pas obligatoire. Ce mode de réalisation permet de conserver un intervalle M relativement grand entre deux trames d'analyse, et donc de limiter le débit de transmission requis, tout en limitant les discontinuités susceptibles d'apparaître en raison de la taille de cet intervalle par rapport aux échelles de temps typiques des variations des paramètres du signal audio, notamment les coefficients cepstraux et la fréquence fondamentale.

Les figures 23 à 25 montrent d'autres formes de réalisation des moyens employés pour traiter les coefficients cepstraux cx_sup délivrés par le module de TFRI 13 de la figure 1, représentant l'enveloppe supérieure.

Dans les trois cas, les modules de post-lifrage 15, de normalisation 16, de quantification 18 et de calcul des amplitudes spectrales 28 sont essentiellement identiques à ceux précédemment décrits en référence à la

figure 1. En outre, des modules de post-lifrage 140, de lissage 141 et d'extraction de phase minimale 142 sont prévus pour traiter les coefficients cepstraux post-liftrés et quantifiés cx_sup_q délivrés par le module de quantification 18. Ces modules 140-142 fonctionnent essentiellement
5 comme les modules correspondants 55-57 du décodeur de la figure 8.

Dans la forme de réalisation montrée sur la figure 23, le module d'adaptation 144 accomplit une fonction semblable à celle du module 29 de la figure 1. Mais l'adaptation n'est pas réalisée sur la seule base du module du spectre. Le module 144 détermine le meilleur jeu de coefficients pour le post-lifre 15 en minimisant l'écart entre le spectre du signal audio, en module $|X|$ et
10 en phase ϕ_X , et des valeurs complexes recalculées pour une ou plusieurs des harmoniques de la fréquence fondamentale. Les modules de ces dernières valeurs complexes sont données par le module de calcul 28, et leurs phases correspondent aux phases minimales $\phi(k)$ fournies par le module d'extraction
15 142. Pour réaliser l'adaptation, le module 144 peut prendre en compte toute distance appropriée dans le plan complexe, par exemple la distance euclidienne.

Ainsi, l'adaptation du post-lifre 15 par le module 144 tient compte de façon combinée d'aspects fréquentiels du signal, reflétés par le module du spectre, et des aspects temporels, reflétés par la phase du spectre.
20

Comme représenté en pointillés sur la figure 23, le post-lifre 140 peut également être adaptatif, l'adaptation effectuée par le module 144 portant conjointement sur les deux post-lifres 15, 140. Dans ce cas, le post-lifre 55 du décodeur (figure 8) est adapté, comme le post-lifre 140, en fonction de
25 paramètres $iLif$ que le module d'adaptation 144 fournit au multiplexeur 6 pour qu'il les inclue dans le flux numérique Φ . Typiquement, quelques jeux de coefficients γ_1, γ_2 sont prévus pour les post-lifres 140 et 55, et le module 144 procède à un test exhaustif de ces différents jeux de coefficients pour retenir celui qui minimise l'écart dans le plan complexe.

30 Dans l'exemple représenté sur la figure 24, le module d'adaptation 29 du post-lifre 15 est identique à celui de la figure 1. La figure 24 montre un module 145 d'estimation d'une courbe de masquage permettant au module 29 de sélectionner, pour la minimisation de l'écart en module, la ou les fréquences harmoniques qui dépassent le plus de la courbe de masquage calculée sur la
35 base du spectre en module $|X|$, comme décrit précédemment.

Le post-litre 140 de la figure 24 est adapté séparément par un module 146 qui réalise la minimisation des écarts entre la phase ϕ_X du spectre du signal et la phase minimale $\phi(k)$ calculée par le module 142 pour une ou plusieurs des harmoniques. Là aussi, les harmoniques sélectionnées pour le calcul de l'écart de phase minimisé peuvent l'être en fonction de la courbe de masquage estimée par le module 145. Le module 146 fournit au multiplexeur de sortie 6 du codeur les paramètres ~~ilif qui représentent le post-litre optimal~~ 140, pour qu'ils soient utilisés dans le post-filtre 55 du décodeur.

Dans l'exemple illustré par la figure 25, le post-litre 140 servant dans le calcul des phases minimales n'est pas adaptatif. Les phases minimales $\phi(k)$ calculées par le module 142 pour les harmoniques de la fréquence fondamentale sont comparées aux phases ϕ_X du spectre du signal audio, et l'écart de phase fait l'objet d'une quantification par un module 148. Les index de quantification correspondants $i\Delta\phi$ sont fournis par le module 148 au multiplexeur de sortie 6 du codeur.

Dans un décodeur (figure 26) correspondant à un codeur selon la figure 25, un module 149 exploite ces index de quantification $i\Delta\phi$ fournis par le démultiplexeur 45 pour obtenir les valeurs des écarts de phase quantifiés, qu'un additionneur 150 ajoute aux phases minimales $\phi(k)$ calculées par le module 57 (les post-litres 140 et 55 étant identiques). Les phases fournies par l'additionneur 150 sont ensuite utilisées par le module 54 qui synthétise les raies spectrales de la composante harmonique X_v .

L'écart de phase quantifié par le module 148, et que les modules 149 et 150 du décodeur utilisent pour corriger les phases minimales $\phi(k)$, peut être de deux natures :

- il peut représenter, pour chaque fréquence d'index i correspondant à une harmonique d'ordre k de la fréquence fondamentale F_0 la différence entre la phase $\phi_X(i)$ du spectre du signal à la fréquence i et la phase minimale $\phi(k)$ calculée par le module 142 pour l'harmonique k ;
- alternativement ou cumulativement, cet écart de phase peut représenter la variation de la phase ϕ_X du spectre sur la largeur d'un ou plusieurs pics spectraux correspondant à des harmoniques du signal, cette variation étant relative à la phase minimale $\phi(k)$ affectée aux pics en question.

Dans les deux cas, le ou les pics pour lesquels l'écart de phase est quantifié peuvent être choisis en fonction de l'énergie spectrale représentée par l'enveloppe supérieure, qui est disponible au codeur et au décodeur, ce qui permet au décodeur de déterminer à quelle raie spectrale les écarts doivent être appliqués.

Dans le premier cas, les écarts de phase peuvent faire l'objet d'une quantification scalaire, ou vectorielle si on les regroupe pour plusieurs pics.

Dans le second cas, la variation de la phase ϕ_x autour de la phase minimale $\phi(k)$ sur la largeur d'un pic harmonique (déterminée par la largeur de la raie de référence utilisée par le module 54), peut être représentée simplement par la pente d'un segment linéaire sélectionné comme étant celui qui présente une distance quadratique minimale avec la courbe de variation de phase du spectre sur la largeur de la raie, et éventuellement par un décalage à l'origine.

Ces pentes peuvent faire l'objet d'une quantification scalaire, ou vectorielle si on les regroupe pour plusieurs pics.

La quantification des variations de phase sur les pics harmoniques peut porter sur l'ensemble des fréquences harmoniques. Une autre possibilité est de quantifier plusieurs pentes obtenues chacune par moyennage des pentes aux harmoniques sur une ou plusieurs sous-bandes du spectre. Ce moyennage peut être pondéré pour prendre en compte les énergies relatives aux différentes fréquences harmoniques, représentées par l'enveloppe supérieure.

Le module 148 peut également modéliser la variation de phase sur la largeur d'un pic par une courbe plus complexe qu'un segment linéaire, par exemple un spline, dont les paramètres sont quantifiés pour être transmis au décodeur.

Une autre possibilité est d'effectuer un apprentissage préalable de modèles de phase aux harmoniques, représentatifs des variations de la phase sur la largeur des pics, observées sur un corpus de signaux de référence. Ces modèles sont rangés dans un dictionnaire mémorisé par les modules 148 et 149. Le module 148 du codeur détermine les index $i\Delta\phi$ correspondant aux adresses des modèles les plus proches des variations de phase au voisinage des pics harmoniques considérés, et le module 149 du décodeur récupère ces modèles pour la synthèse de la phase de la composante harmonique.

REVENDICATIONS

1. Procédé de codage d'un signal audio (x), dans lequel on estime une fréquence fondamentale (F_0) du signal audio, on détermine un spectre du signal audio par une transformée dans le domaine fréquentiel d'une trame du signal audio, et on inclut dans un flux numérique de sortie (Φ) des données de
5 codage d'une composante harmonique du signal audio, comprenant des données représentatives d'amplitudes spectrales associées à des fréquences multiples de la fréquence fondamentale, caractérisé en ce que l'amplitude spectrale associée à une desdites fréquences multiples de la fréquence
10 fondamentale est un maximum local du module du spectre au voisinage de ladite fréquence multiple.
2. Procédé selon la revendication 1, dans lequel lesdites données représentatives d'amplitudes spectrales associées à des fréquences multiples de la fréquence fondamentale (F_0) sont obtenues au moyen de coefficients
15 cepstraux (cx_sup) calculés en transformant dans le domaine cepstral une enveloppe supérieure comprimée (LX_sup) du spectre du signal audio.
3. Procédé selon la revendication 2, dans lequel l'enveloppe supérieure comprimée (LX_sup) est déterminée par interpolation desdites amplitudes spectrales associées aux fréquences multiples de la fréquence
20 fondamentale (F_0) avec application d'une fonction de compression spectrale.
4. Procédé selon la revendication 3, dans lequel l'interpolation est effectuée entre des points dont l'abscisse est une fréquence multiple de la fréquence fondamentale (F_0) et dont l'ordonnée est l'amplitude spectrale associée à ladite fréquence multiple, comprimée ou non comprimée.
- 25 5. Procédé selon l'une quelconque des revendications 2 à 4, dans lequel la transformation dans le domaine cepstral de l'enveloppe supérieure comprimée (LX_sup) est effectuée suivant une échelle de fréquences non-linéaire.
6. Procédé selon l'une quelconque des revendications 2 à 5, dans
30 lequel les coefficients cepstraux (cx_sup) sont quantifiés pour former lesdites

données représentatives des amplitudes spectrales associées aux fréquences multiples de la fréquence fondamentale (F_0).

7. Procédé selon la revendication 6, dans lequel la quantification des coefficients cepstraux (cx_sup) porte sur un résidu de prédiction pour chacun
5 des coefficients cepstraux.

- ~~8. Procédé selon la revendication 7, dans lequel le résidu de prédiction~~
pour un coefficient cepstral est de la forme $(cx[n,i] - \alpha(i).rcx_q[n-1,i]) / [2 - \alpha(i)]$,
où $cx[n,i]$ désigne une valeur courante dudit coefficient cepstral, $rcx_q[n-1,i]$
désigne une valeur précédente du résidu de prédiction quantifié, et $\alpha(i)$ désigne
10 un coefficient de prédiction.

9. Procédé selon la revendication 7 ou 8, dans lequel on emploie des prédicteurs différents pour déterminer les résidus de prédiction pour au moins deux des coefficients cepstraux.

10. Procédé selon l'une quelconque des revendications 6 à 9, dans
15 lequel les coefficients cepstraux (cx_sup) sont répartis en plusieurs sous-vecteurs cepstraux quantifiés séparément par une quantification vectorielle portant sur un résidu de prédiction des coefficients cepstraux.

11. Procédé selon l'une quelconque des revendications 6 à 10, dans lequel on normalise les coefficients cepstraux (cx_sup) avant la quantification, en modifiant le coefficient cepstral d'ordre 0 de telle sorte que l'amplitude spectrale associée à une fréquence multiple de la fréquence fondamentale (F_0) soit représentée de manière exacte par les coefficients cepstraux normalisés.
20

12. Procédé selon l'une quelconque des revendications 6 à 11, dans lequel les coefficients cepstraux (cx_sup) sont transformés par liftrage dans le
25 domaine cepstral avant d'être quantifiés.

13. Procédé selon la revendication 12, dans lequel le liftrage est de la forme $c_p(i) = [1 + \gamma_2^i - \gamma_1^i].c(i) - (\mu^i/i)$, où $c_p(i)$ et $c(i)$ désignent le coefficient cepstral d'ordre $i > 0$ respectivement avant et après liftrage, γ_1 et γ_2 sont des coefficients compris entre 0 et 1, et μ est un coefficient de préaccentuation.

14. Procédé selon la revendication 13, dans lequel $\mu = (\gamma_2 - \gamma_1).c(1)$.
15. Procédé selon l'une quelconque des revendications 12 à 14, dans lequel on recalcule une valeur du module du spectre du signal audio à au moins une fréquence multiple de la fréquence fondamentale (F_0) sur la base des coefficients cepstraux transformés et quantifiés (cx_sup_q), et on adapte ledit liftrage de façon à minimiser un écart de module entre le spectre du signal audio et au moins une valeur de module recalculée.
16. Procédé selon l'une quelconque des revendications 12 à 14, dans lequel on recalcule une valeur du module du spectre du signal audio à au moins une fréquence multiple de la fréquence fondamentale (F_0) sur la base des coefficients cepstraux transformés et quantifiés (cx_sup_q), on retransforme les coefficients cepstraux par liftrage et lissage dans le domaine cepstral, on calcule des phases minimales ($\phi(k)$) du signal audio à des fréquences multiples de la fréquence fondamentale sur la base des coefficients cepstraux retransformés ($cxl[n]$), et on adapte le liftrage effectué avant la quantification de façon à minimiser un écart entre le spectre du signal audio et au moins une valeur complexe dont le module a une valeur recalculée pour une fréquence multiple de la fréquence fondamentale et dont la phase est donnée par la phase minimale calculée pour ladite fréquence multiple.
17. Procédé selon la revendication 16, dans lequel les liftrages effectués avant et après la quantification sont adaptés conjointement de façon à minimiser ledit écart, et dans lequel on inclut dans les données de codage de la composante harmonique des paramètres (iLif) représentatifs du liftrage adapté effectué après la quantification.
18. Procédé selon l'une quelconque des revendications 15 à 17, dans lequel l'écart minimisé pour l'adaptation du liftrage se rapporte à au moins une fréquence multiple de la fréquence fondamentale (F_0), sélectionnée sur la base de l'importance du module du spectre en valeur absolue.
19. Procédé selon l'une quelconque des revendications 15 à 17, dans lequel on estime une courbe de masquage spectral du signal audio au moyen d'un modèle psycho-acoustique, et l'écart minimisé pour l'adaptation du liftrage se rapporte à au moins une fréquence multiple de la fréquence fondamentale

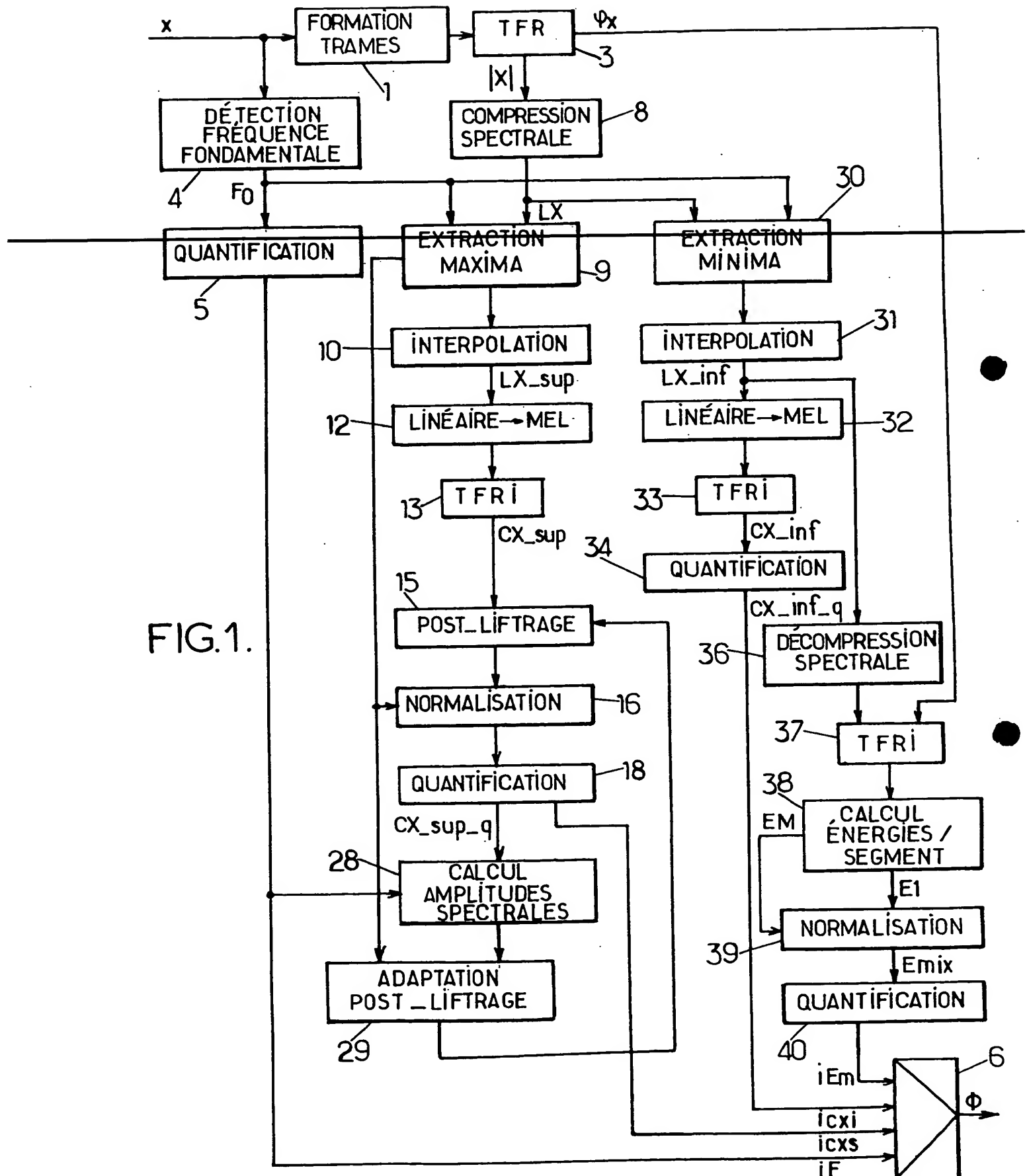
(F_0), sélectionnée sur la base de l'importance du module du spectre relativement à la courbe de masquage.

20. Procédé selon la revendication 2, dans lequel le spectre du signal audio et les coefficients cepstraux (cx_sup) résultant de la transformation de l'enveloppe supérieure comprimée sont déterminés pour des trames successives de N échantillons du signal audio qui présentent des recouvrements mutuels, et dans lequel lesdites données représentatives d'amplitudes spectrales associées aux fréquences multiples de la fréquence fondamentale estimée (F_0), obtenues au moyen des coefficients cepstraux calculés en transformant l'enveloppe supérieure comprimée, sont incluses dans le flux numérique de sortie (Φ) pour seulement un sous-ensemble des trames.

21. Procédé selon la revendication 20, dans lequel, pour les trames ne faisant pas partie dudit sous-ensemble, on inclut dans le flux numérique de sortie (Φ) des données ($icx[n-1/2]$) de quantification d'une erreur d'interpolation ($ecx[n-1/2]$) des coefficients cepstraux résultant de la transformation de l'enveloppe supérieure comprimée (LX_sup).

22. Procédé selon la revendication 20, dans lequel, pour les trames ne faisant pas partie dudit sous-ensemble, on détermine un filtre interpolateur optimal (128) pour les coefficients cepstraux résultant de la transformation de l'enveloppe supérieure comprimée (LX_sup), et on inclut dans le flux numérique de sortie (Φ) des données (iP) représentant ledit filtre interpolateur optimal.

23. Codeur audio, comprenant des moyens pour exécuter un procédé selon l'une quelconque des revendications 1 à 22.



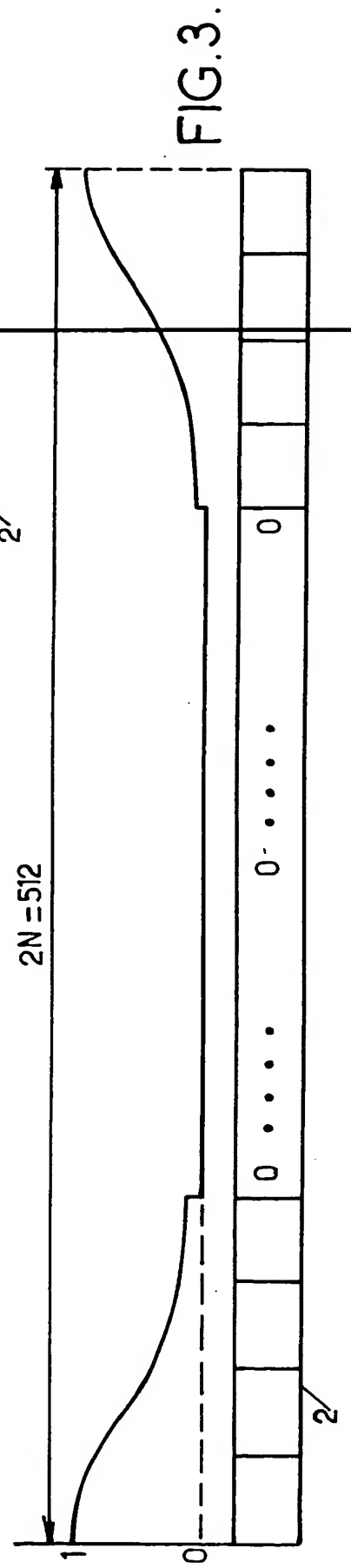
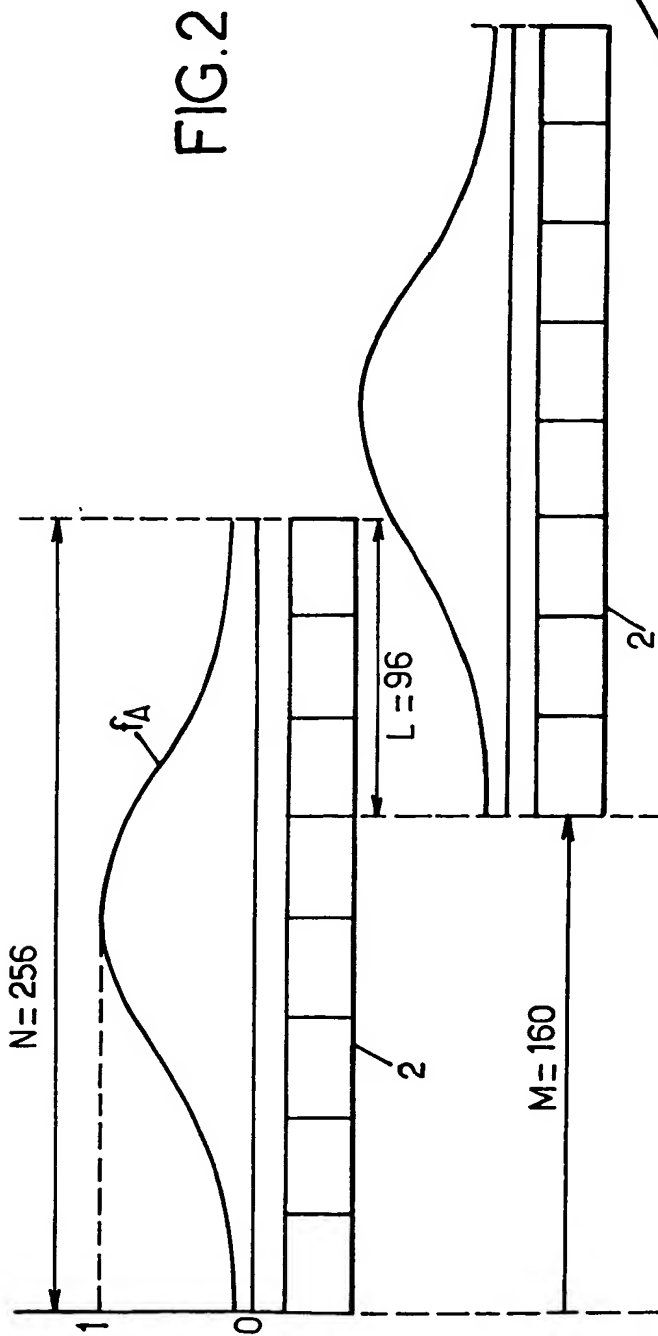


FIG.4.

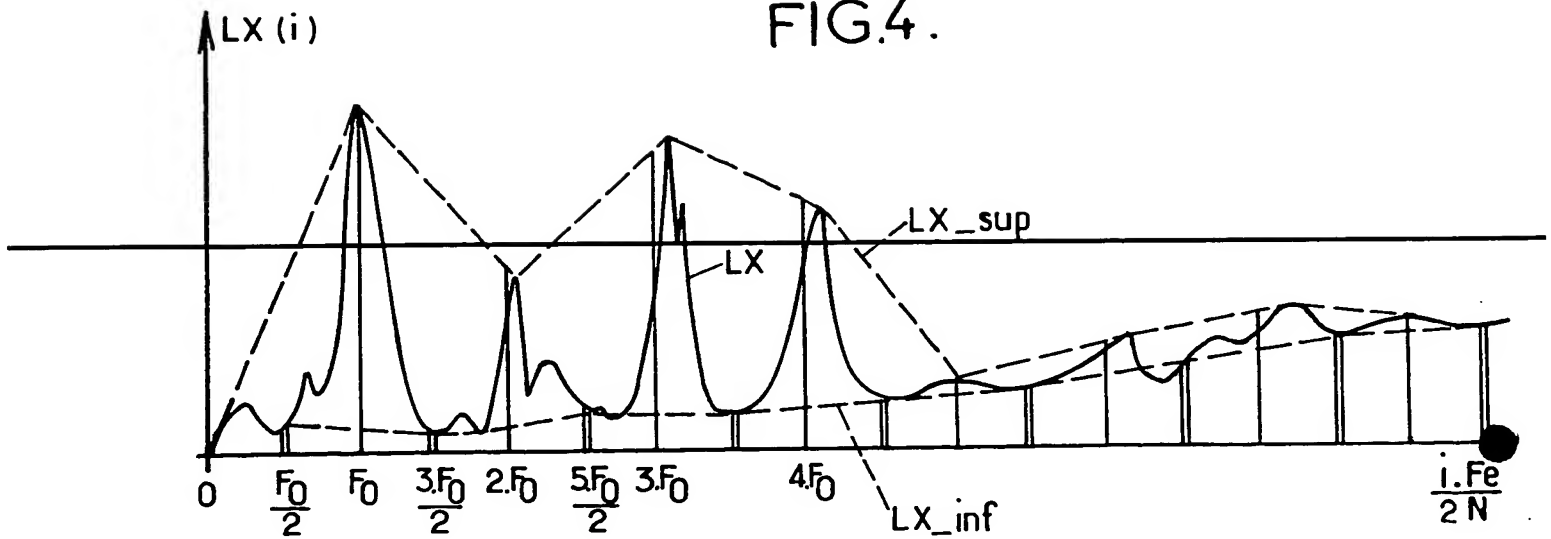


FIG.5.

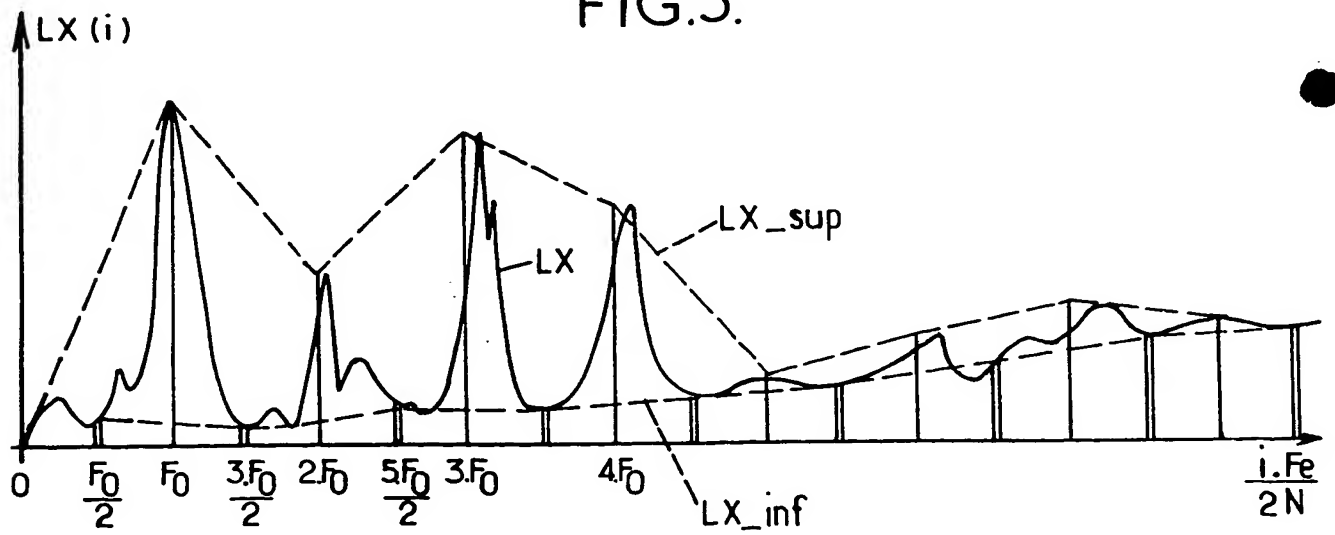


FIG.6.

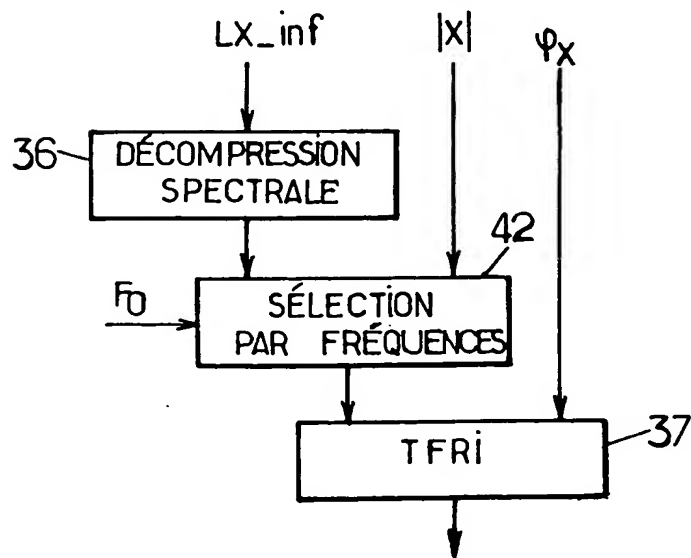
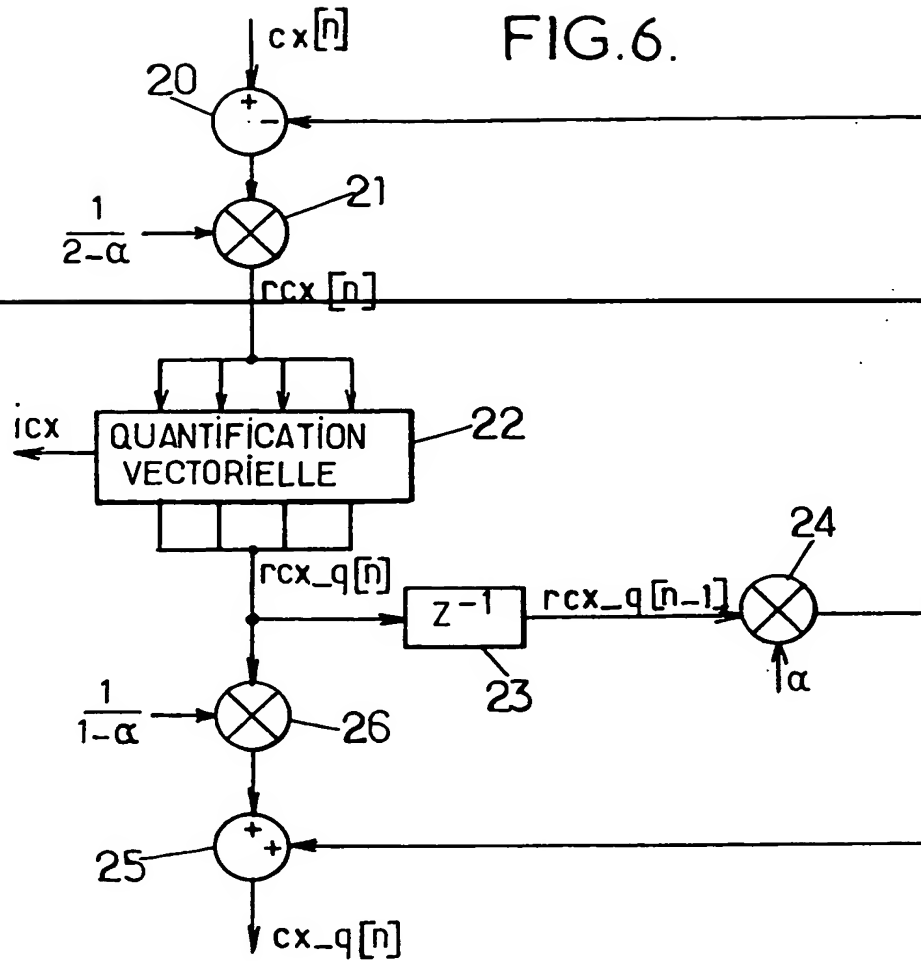


FIG.7.

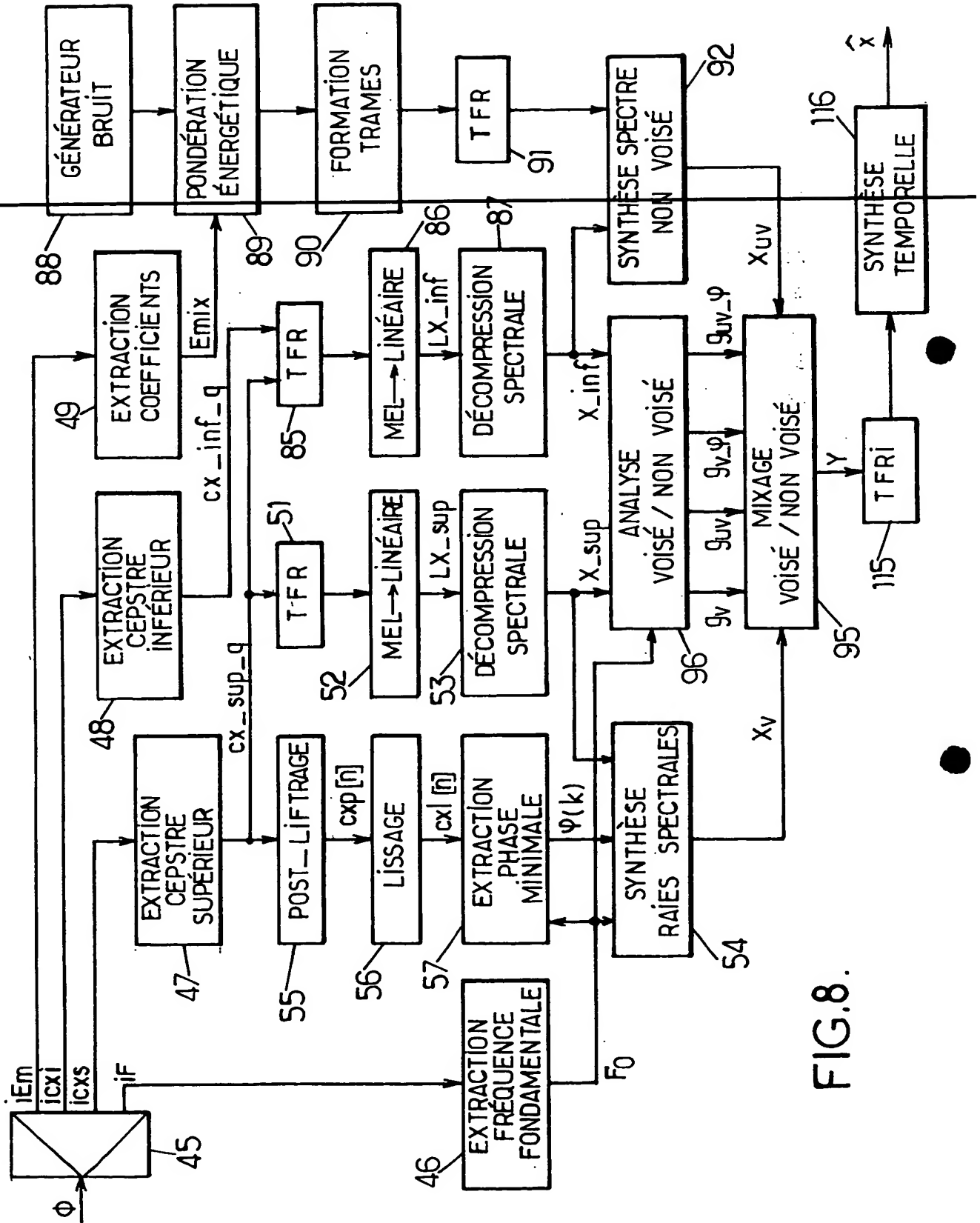
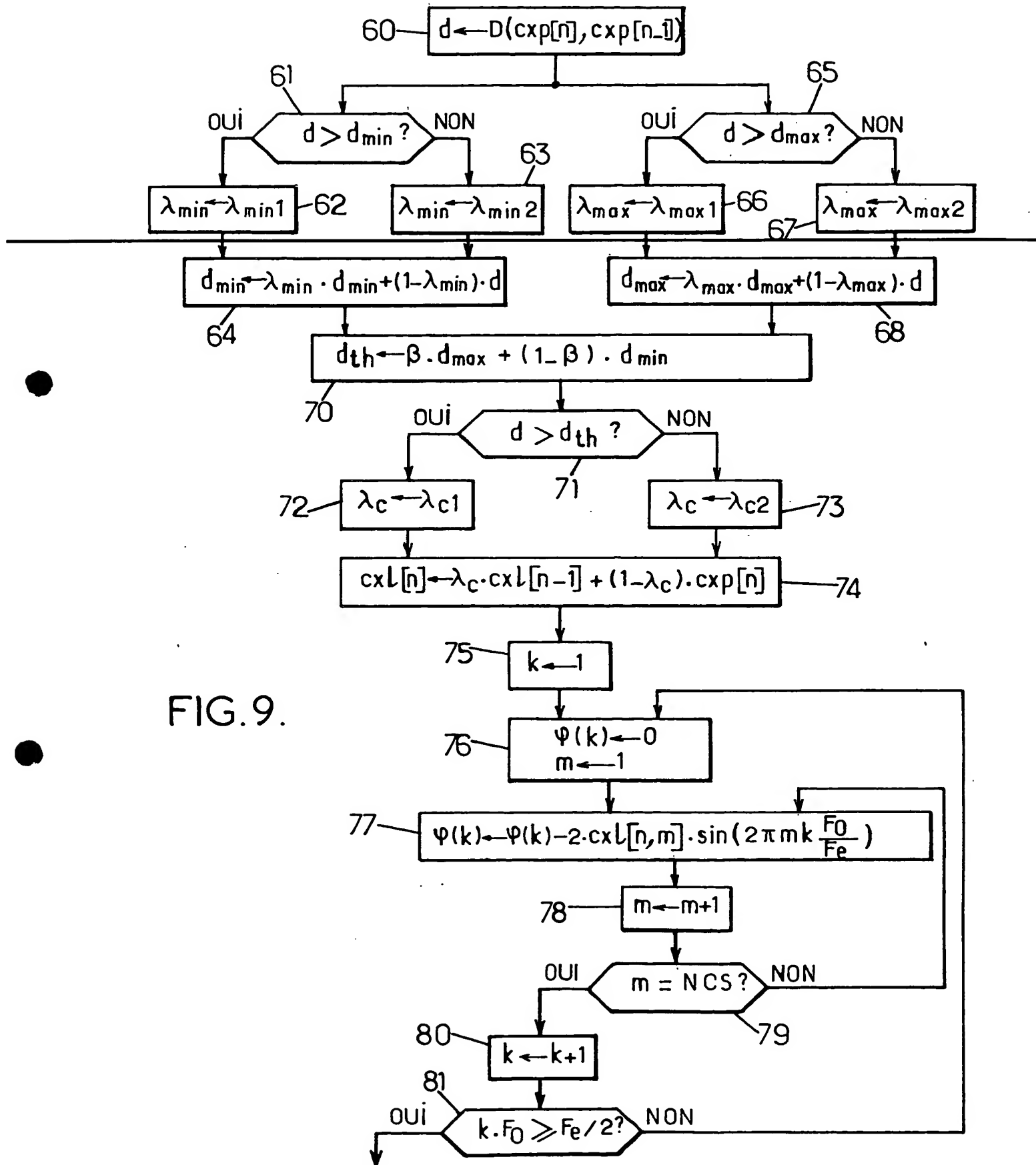


FIG. 8.



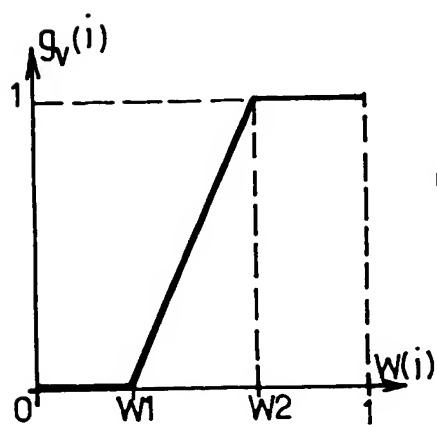
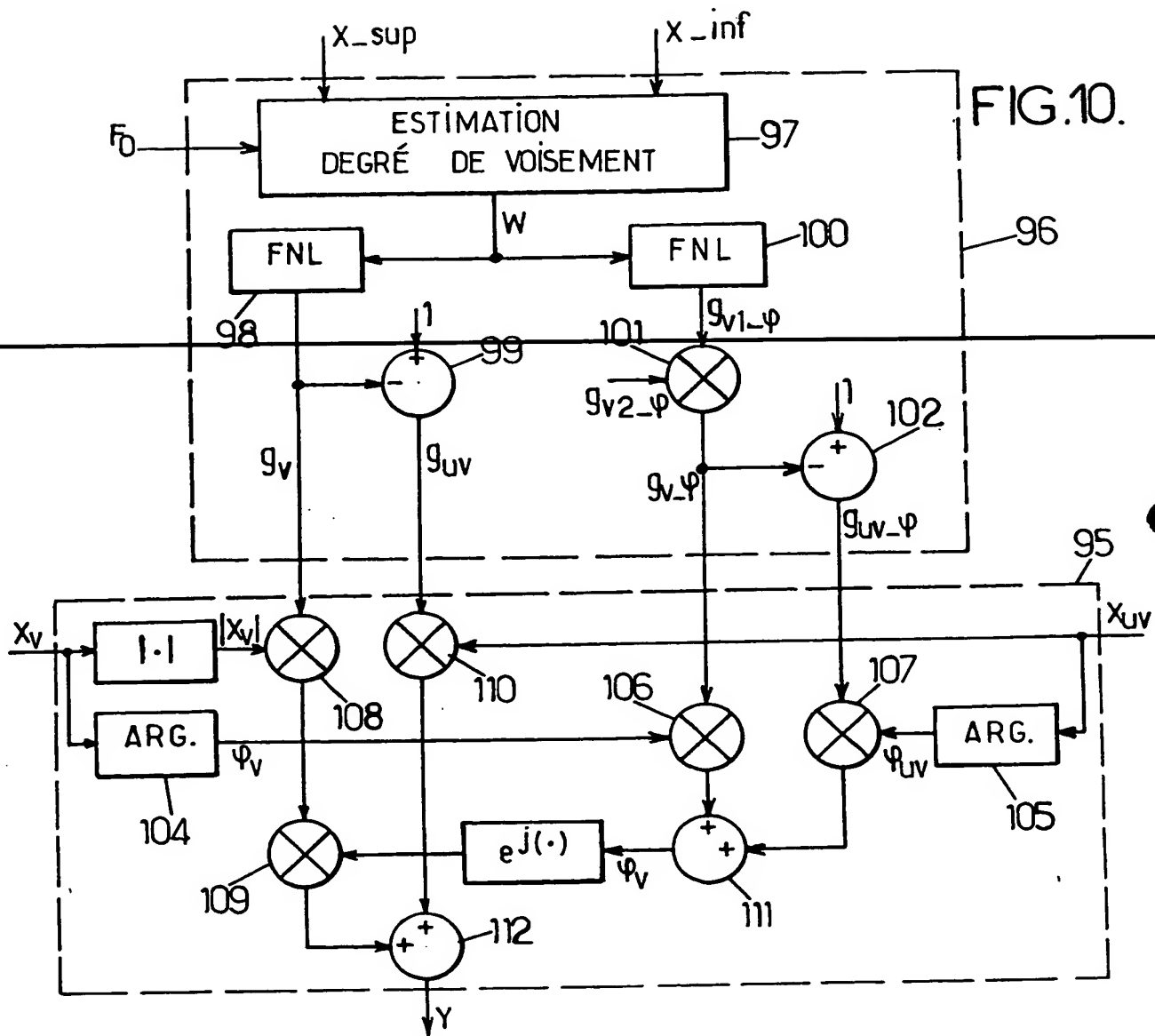


FIG.11

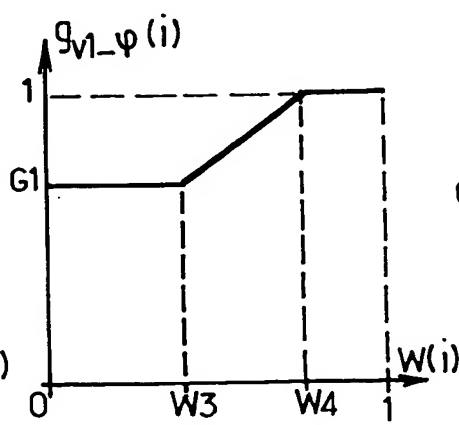


FIG.12.

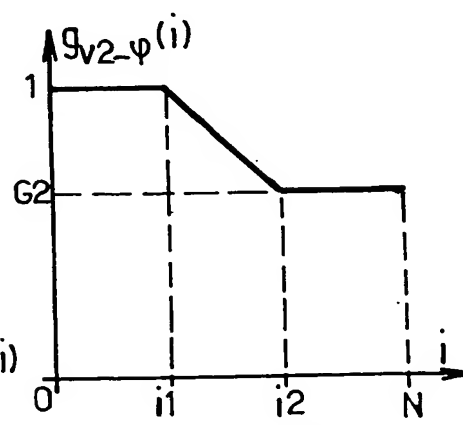


FIG.13.

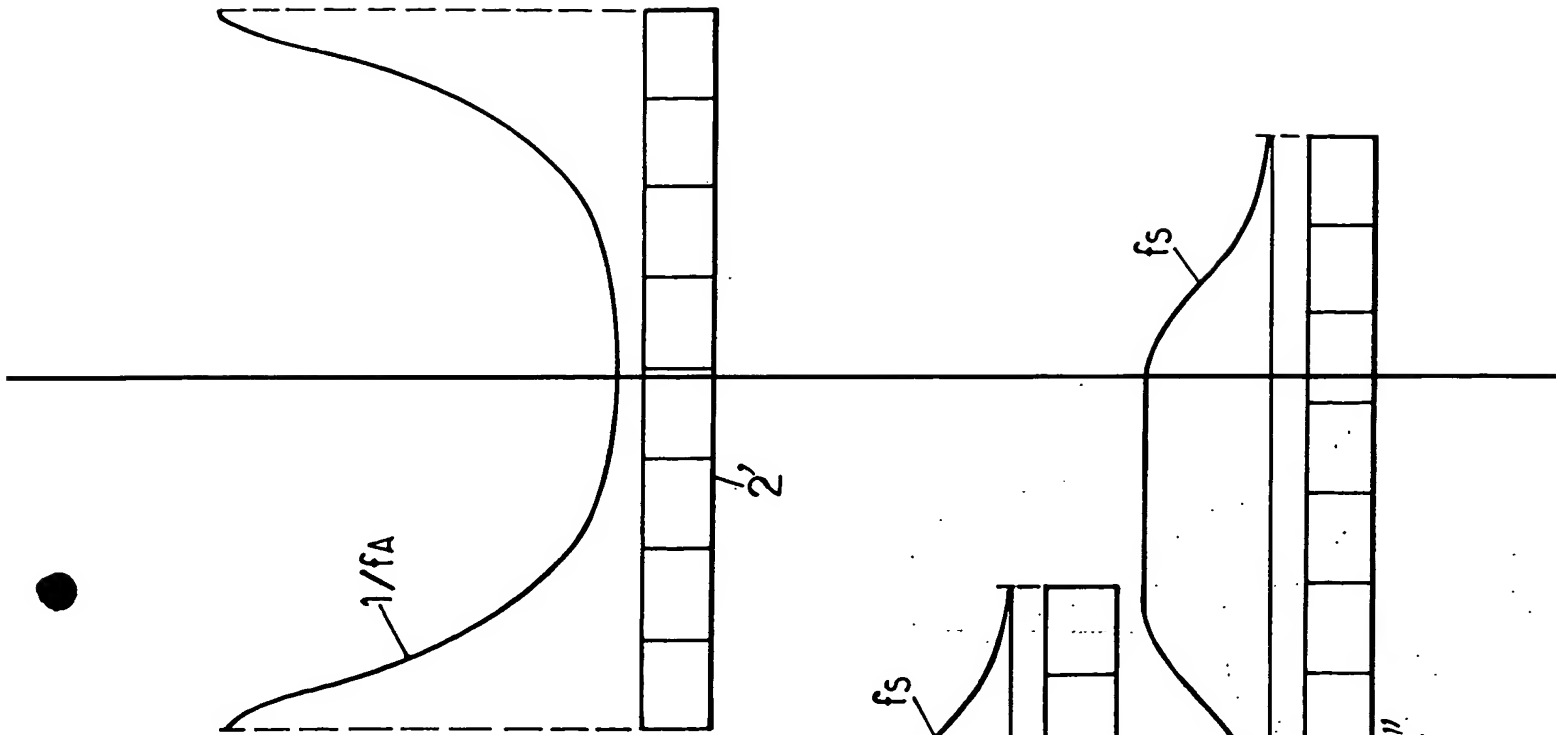


FIG. 14.

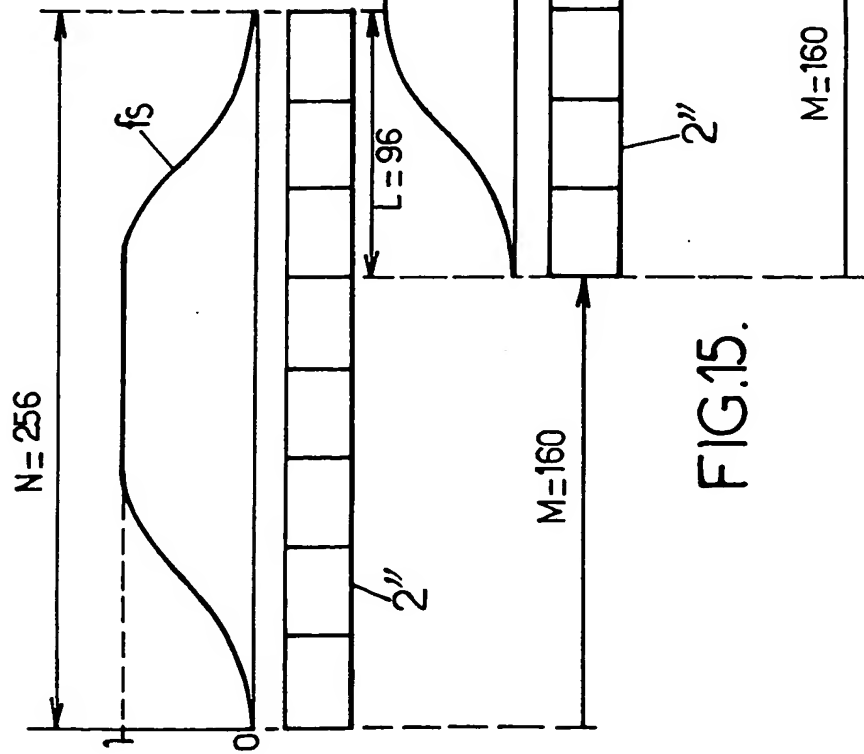


FIG. 15.

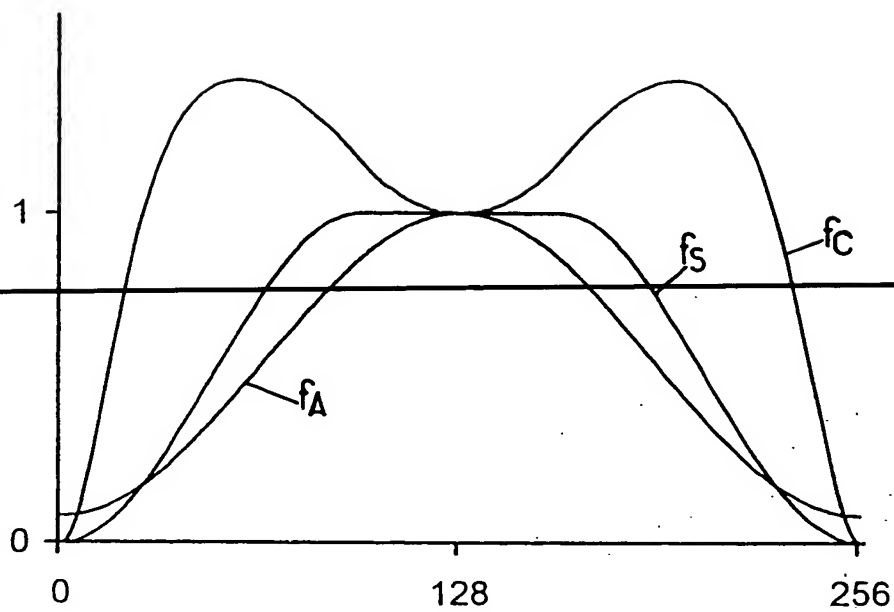


FIG.16.

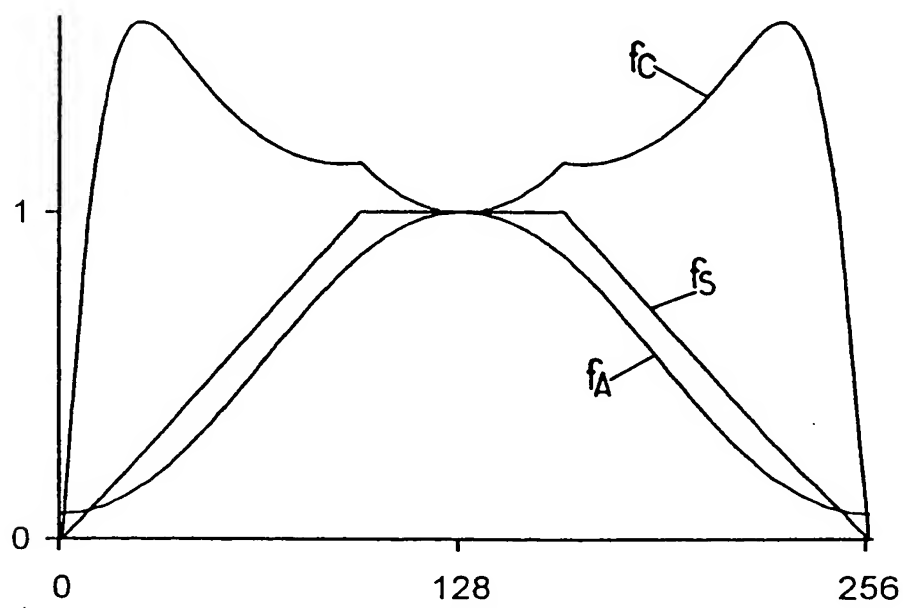


FIG.17.

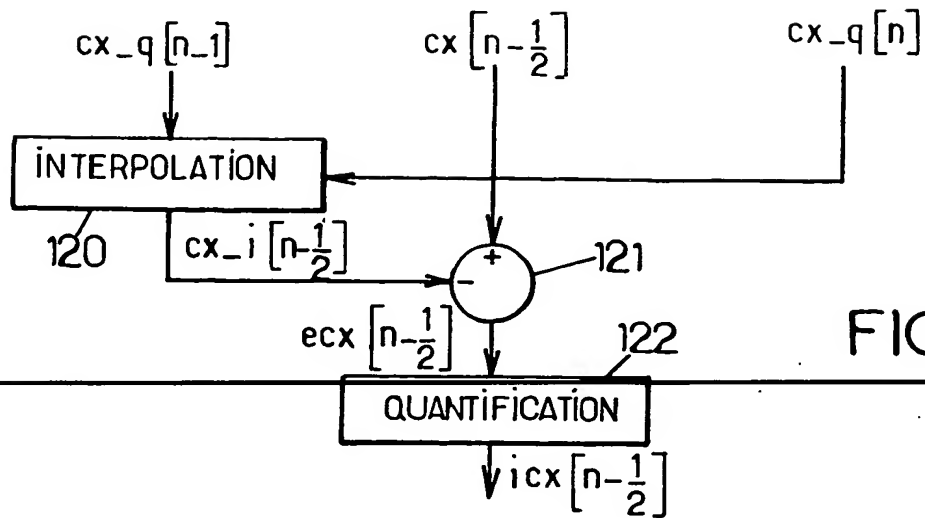


FIG.18.

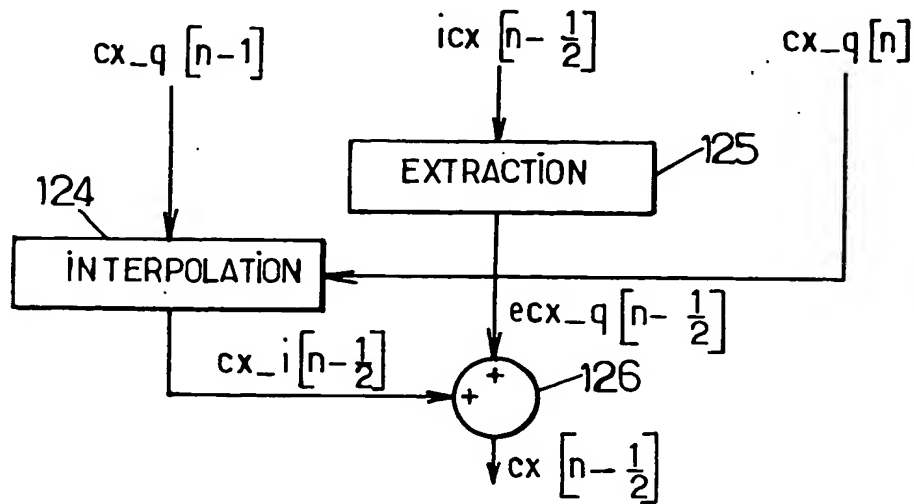


FIG.19.

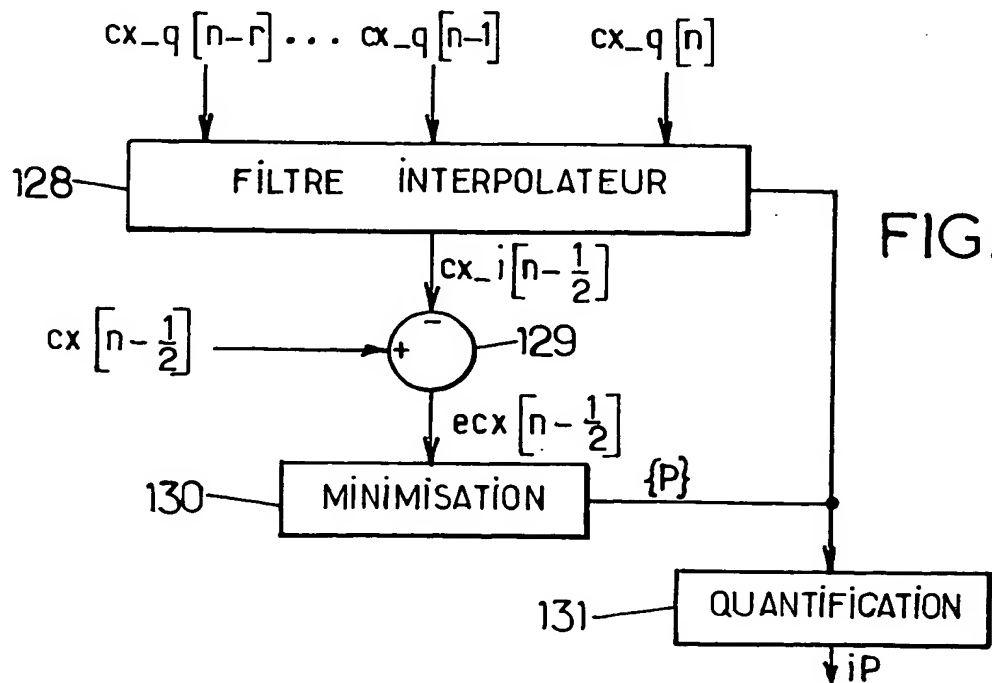


FIG.20.

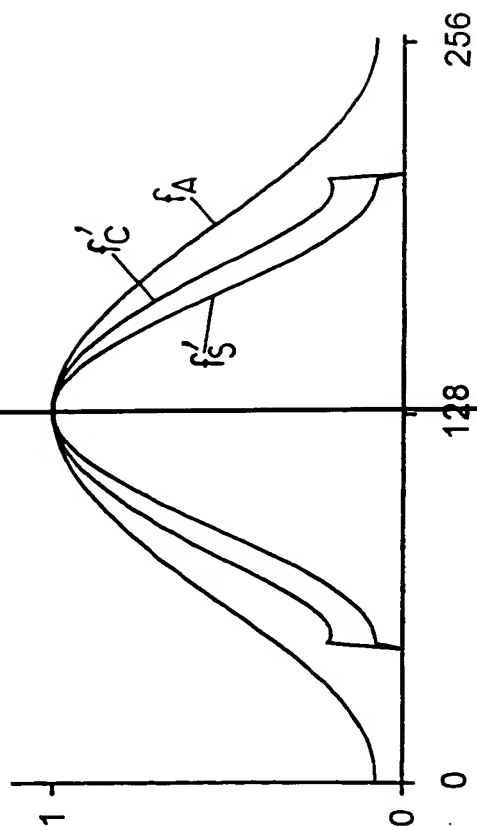


FIG.22.

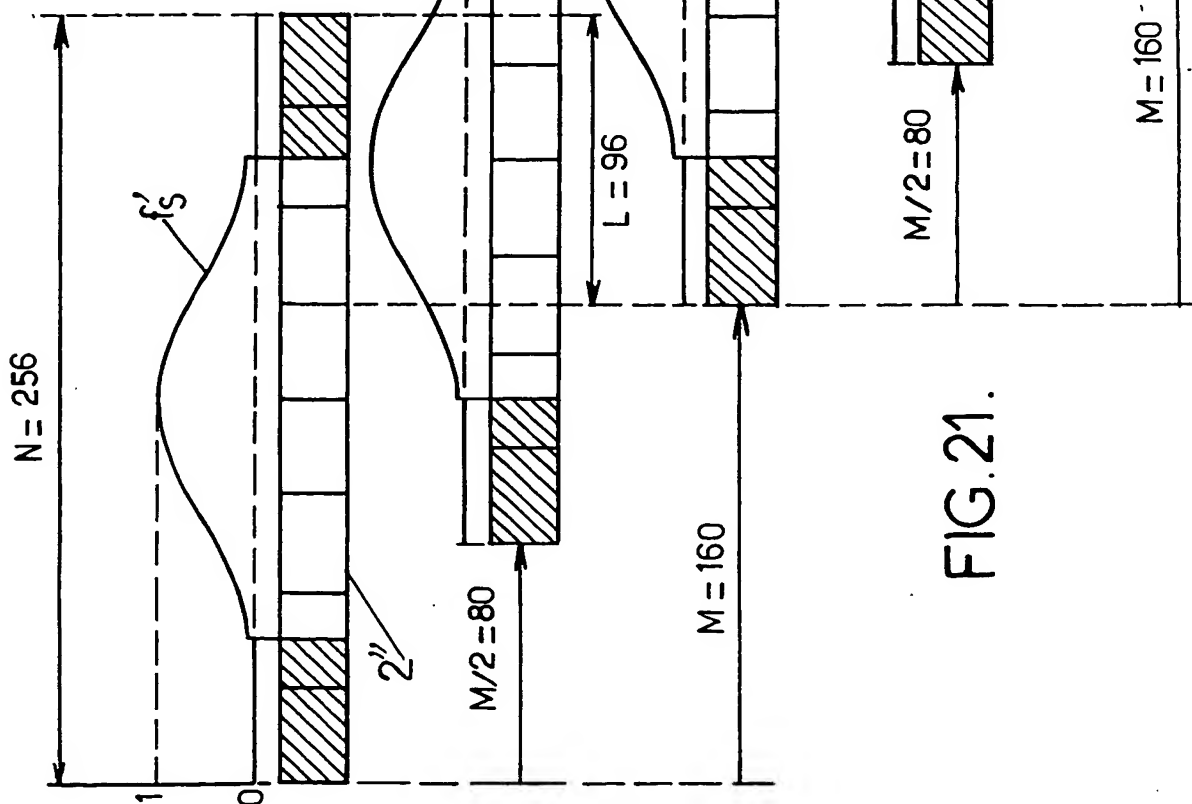


FIG.21.

FIG.23.

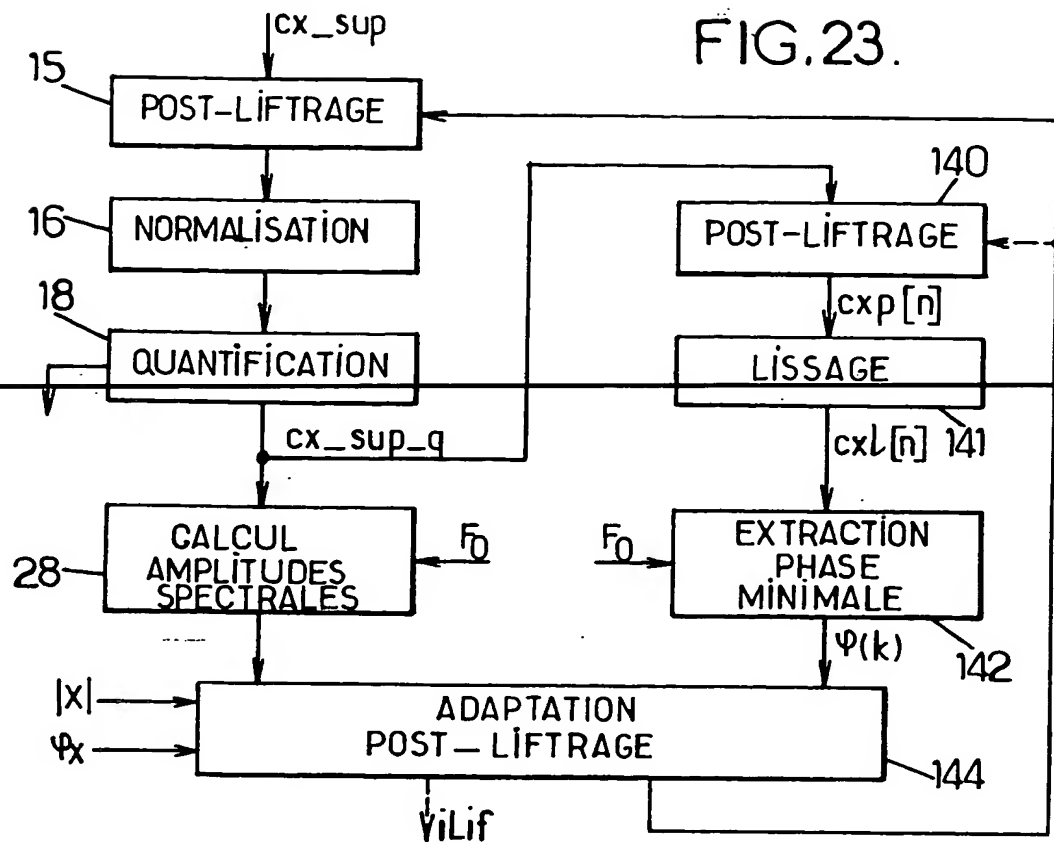


FIG.24.

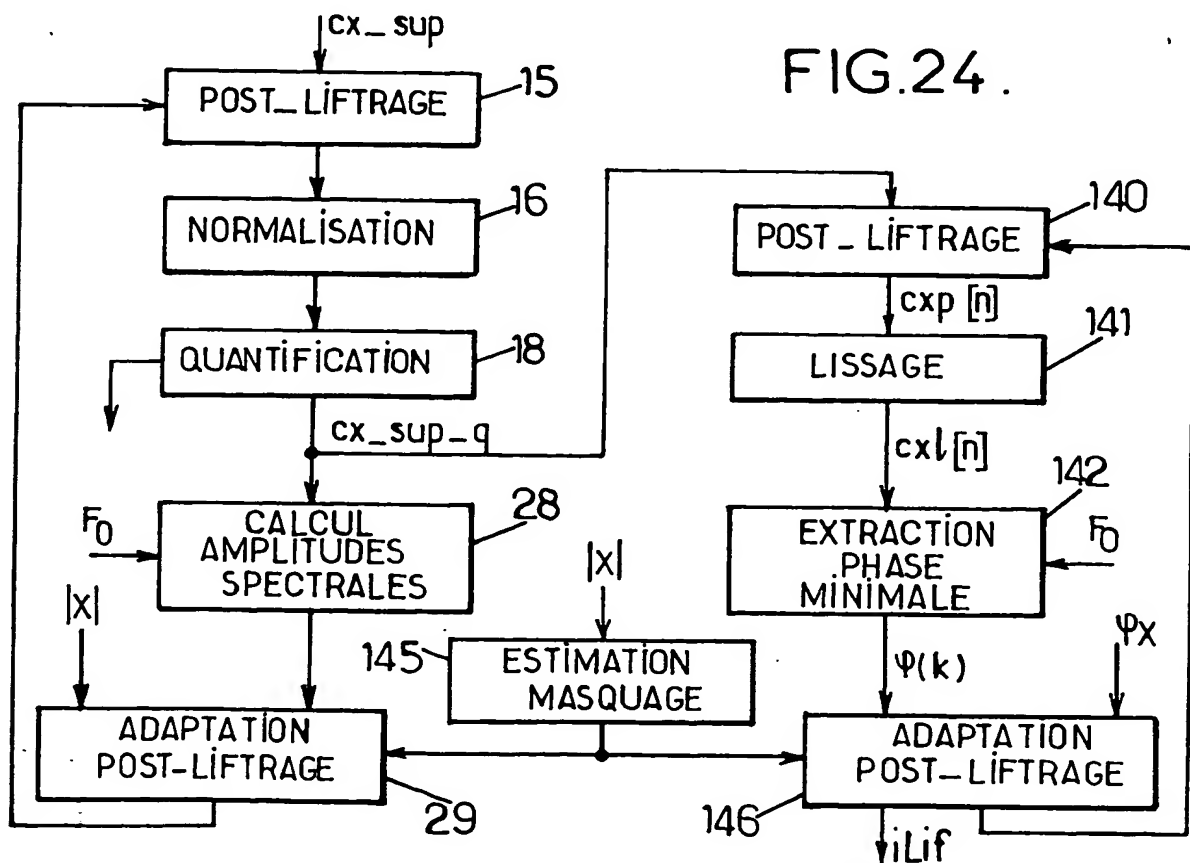


FIG.25.

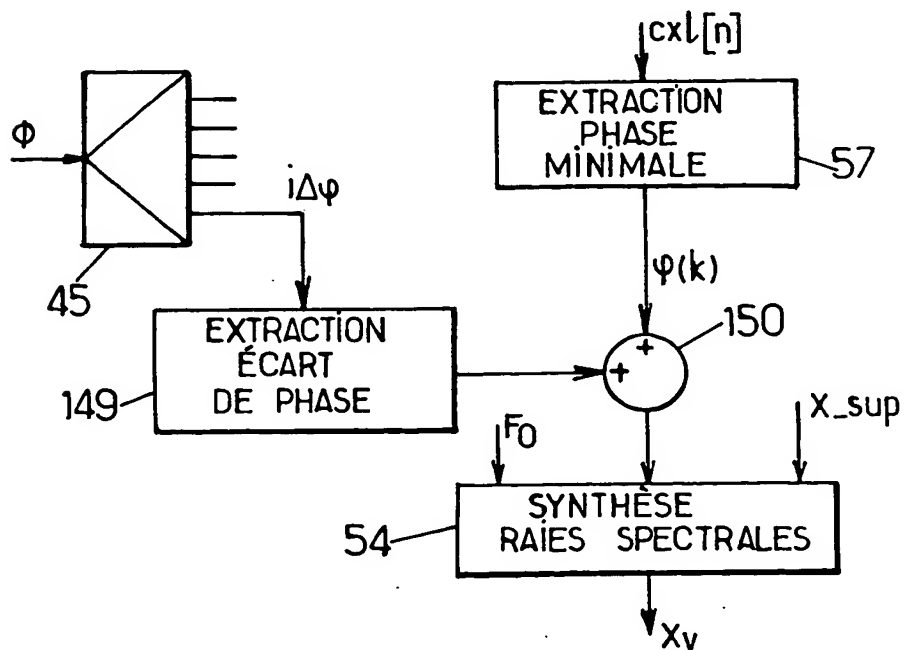
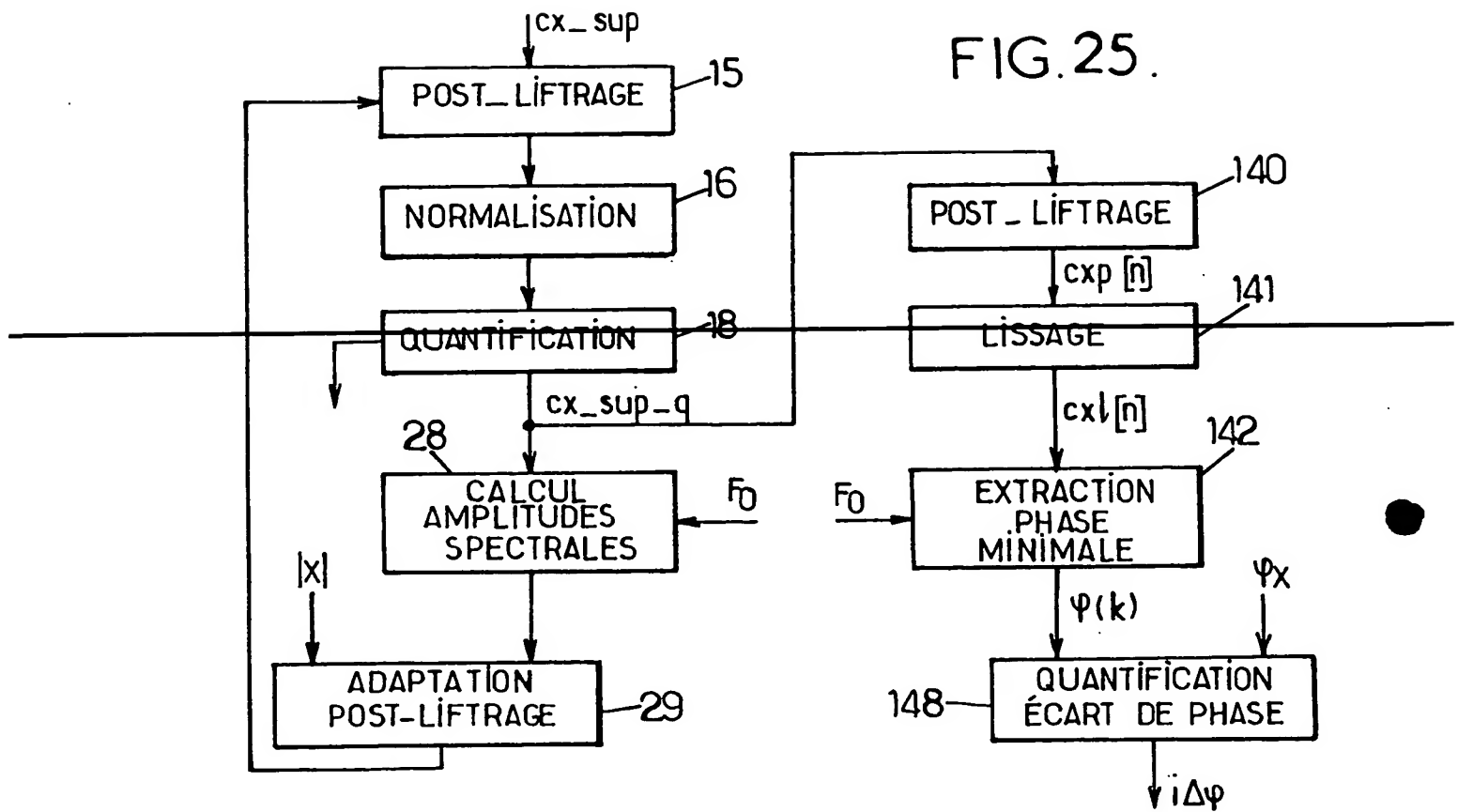


FIG.26.

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER: _____**

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)